Universidade Federal
do Rio de Janeiro
Escola Politécnica

Classificação dos estados de evolução microstrutural de tubos de reforma a vapor de diferentes composições químicas utilizando inteligência artificial – uma abordagem de END 4.0

Pedro Enrique Monforte Brandão Marques

Projeto de graduação apresentado ao Curso de Engenharia de Materiais da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro de Materiais.

Orientador: Gabriela Ribeiro Pereira
Coorientador: Clara Johanna Pacheco

Rio de Janeiro

Novembro/2020

Classificação dos estados de evolução microstrutural de tubos de reforma a vapor de diferentes composições químicas utilizando inteligência artificial – uma abordagem de END 4.0

Pedro Enrique Monforte Brandão Marques

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA DE MATERIAIS DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE MATERIAIS.

Examinado por:

_____
Gabriela Ribeiro Pereira, D.Sc.

_____
Adriana da Cunha Rocha, D.Sc.

_____
Luiz Henrique Almeida, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2020

Marques, Pedro Enrique Monforte Brandão

Classificação dos estados de evolução microstrutural de tubos de reforma a vapor de diferentes composições químicas utilizando inteligência artificial – uma abordagem de END 4.0 / Pedro Enrique Monforte Brandão Marques - Rio de Janeiro: UFRJ / Escola Politécnica, 2020.

VII, 57 p.: il.; 29,7 cm.

Orientador: Gabriela Ribeiro Pereira

Projeto de graduação – UFRJ / Escola Politécnica / Curso de Engenharia de Materiais, 2020.

Referências Bibliográficas: p. 55 - 57

1. Ensaios não Destrutivos 2. Inteligência Artificial 3. Estatística aplicada. I. Pareira, Gabriela Ribeiro. II. Universidade Federal do Rio de Janeiro, Escola Politécnica, Cursop de Engenharia de Materiais. III. Classificação dos estados de evolução microstrutural de tubos de reforma a vapor de diferentes composições químicas utilizando inteligência artificial – uma abordagem de END 4.0.

DEDICATÓRIA

Ao povo brasileiro,
que me deu os meios e motivos para
me tornar um engenheiro.

AGRADECIMENTO

Agradeço primeiramente à minha família, que foi o firmamento sobre o qual toda a minha vida se sustenta. Agradeço à minha mãe Elenice, a pessoa mais forte e corajosa que já conheci, que, apesar de todas as dificuldades e intempéries do mundo que se impuseram a nós, assumiu na sua vida a responsabilidade de levar seus filhos a um tortuoso, e por vezes improvável, caminho até a felicidade. Um caminho que passa por veredas que ela mesma se impediu de trilhar pra permitir que nós pudéssemos percorrê-las. Um caminho que de passos em passos encontramos um espinho, mas de léguas em léguas, uma flor. Agradeço à minha irmã Carolina, que, apesar de nos oferecer desafios diários e aparentemente intermináveis, a cada dia nos faz acreditar mais e mais que tudo vale a pena com doses infindáveis de amor. Tanto amor sincero e puro que prova acima de qualquer suspeita que esse ódio e crueldade tão profundos que parecem enraizados nas estruturas da humanidade são apenas uma fase derrotada e com tempos contados na nossa história. Agradeço à minha avó Shirley e minha tia Thaisa que sempre estiveram do nosso lado com carinho e procupação nos piores e melhores momentos, apoiando e aguardando ansiosamente a conclsuão desse trabalho e desse curso. Agradeço com longa e doída saudade ao meu pai, que apesar de não poder estar presente, sem dúvida alguma é a pessoa mais feliz por eu ter alcançado esse novo estágio da vida. Eu consigo ver o sorriso sem tamanho que ele manteria por semanas por ver o filho se formando, mesmo que não possa mais vê-lo ao meu lado. Jaci permanece brilhando mesmo nas noites escuras.

Agradeço à minha amada companheira Andressa, que entrou na minha vida como um raio de sol em meio à chuva, que aquece e conforta sem queimar, e refresca sem nunca deixar minha alma fria. Esse trabalho, sem sua parceria, sem seu amor e seu carinho teria sido tanto impossível quanto insignificante.

Agradeço a toda a equipe do LNDC, que me ensinaram que a sala de aula é muito maior do que o bloco F! Cesar, ambos os Lucas, Tibério, Iane, Spock, ambas as Anas, Vitor, Adriano, Daniel, Thiago, e muitas e muitos outros que fizeram esse trabalho ser o mais divertido e completo possível. Agradeço especialmente a Johanna, que coorientou esse trabalho e que não poucas vezes dedicou horas de conversa com os meus inúmeros desabafos. As coisas que aprendi nessas conversas não estão, em nem poderiam estar, em nenhum livro.

Agradeço a minha incrível orientadora e amiga Gabriela, que se tornou pra mim não apenas um exemplo acadêmico, mas principalmente um exemplo humano, pedagógico e de coragem frente ao absurdo. Uma professora e engenheira (sim, nada

de física!) comprometida com a educação e a pesquisa libertadoras, insubmissas e populares. Uma profissional que bravamente decidiu estar do lado certo da história.

Agradeço a minha incansável orientadora e amiga Adriana que, apesar de não estar listada como orientadora desse trabalho, me orientou em literalmente tudo que eu fiz, pensei e propus desde que pisei na UFRJ. Ela esteve comigo, me apoiou e teve um papel crucial no cine-debate Metalmat, na criação do Coletivo Minerva Tupiniquim, na refundação da Confraria Acadêmica da Metalmat (CAMM), na minha primeira iniciação científica na graduação, na primeira vez que entrei no LNDC, no primeiro trabalho acadêmico que escrevi e apresentei, no primeiro congresso de engenharia que fui e apresentei trabalho, na porrada do primeiro artigo negado, no primeiro estágio que consegui, enfim, a lista não acaba. Em cada momento difícil e em cada vitória, ela esteve do meu lado, fazendo sempre tudo o que pode para ajudar. Me ensinou que a verdadeira "cachanga real" não se conquista com autoridade, e sim com inteligência e respeito. Não existe uma forma mais clara de provar e exemplificar o papel de uma professora. Ensinar é trilhar um longo caminho de duas direções.

Agradeço aos camaradas das Brigadas Populares, que seguram comigo a bandeira sob a qual decidi dedicar minha vida e minha profissão: cumprir o destino histórico do Brasil, uma nação insubmissa que não desiste do sol da liberdade não importa quão longa seja a noite de sofrimento.

Agradeço aos moradores da Ocupação Povo Sem Medo, principalmente às incansáveis companheiras Leila e Andreia. Foi no chão desse local e dessa luta que pela primeira vez eu me senti um engenheiro, e que entendi o verdadeiro significado dessa palavra. A noite do dia 12 de outubro de 2017 foi a minha verdadeira formatura, muito mais profunda que qualquer formalidade.

Agradeço aos amigos e camaradas brigadistas que me ajudaram com braços e mentes a criar e hoje dar vida ao Coletivo Força Motriz. Foi no nascimento e na infância dele que eu finalmente aprendi a amar e valorizar a profissão que escolhi. Foi a cada desafio árduo que ele nos ofereceu que ficou clara a poderosa ferramenta de luta e de resistência que a engenharia pode ser nas mãos e corações certos. Foi desenhando esse caminho que eu descobri que ele valia a pena.

Agradeço as companheiras e companheiros do Sindicato dos Engenheiros do Rio de Janeiro, que me mostraram que eu não estava sozinho na luta por uma engenharia popular, profudamente comprometida com a libertação do Brasil.

Por último, humildemente agradeço à Nossa Senhora de Aparecida, que escolheu descer dos céus para redimir cada uma das muitas mortes que nós humanos enfrentamos durante a vida. Aquela que decidiu ser a mãe dos pobres e desesperados, que nada tem senão a esperança. Obrigado por nunca nos abandonar.

Resumo do Projeto de Graduação apresentado à Escola Politécnica / UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Materiais.

Classificação dos estados de evolução microstrutural de tubos de reforma a vapor de diferentes composições químicas utilizando inteligência artificial – uma abordagem de END 4.0

Pedro Enrique Monforte Brandão Marques

Novembro/2020

Orientador: Gabriela Ribeiro Pereira
Coorientador: Clara Johanna Pacheco

Curso: Engenharia de Materiais

Aços inoxidáveis austeníticos resistentes ao calor da classe HP são comumente aplicados em condições operacionais extremas, sofrendo um processo de envelhecimento. Esse processo, em condições anormais de operação, pode ser acelerado, podendo gerar a falha por fluência do material. Uma das aplicações desse material é em tubos de fornos de reforma a vapor. Estudos anteriores propuseram métodos de inspeção e classificação não destrutiva dos estados de evolução microstrutural de tubos de fornos de reforma desse material com composições químicas específicas. Nesse trabalho, propõem-se uma nova metodologia de classificação de dados da inspeção de tubos com composições químicas variadas, utilizando técnicas de inteligência artificial e estatística aplicada baseadas na abordagem de END 4.0. Os resultados do uso da metodologia proposta para treinar um modelo único para tubos com duas composições químicas diferentes mostraram que foi possível atingir uma acurácia de 91.8% para a classificação dos dados analisados.

Palavras-chave: Metalurgia, Aços Inoxidáveis Austeníticos, Ensaios Não Destrutivos, Inteligência Artificial, Análise de dados.

Abstract of the Undergraduate Project presented to the POLI / UFRJ as a partial fulfillment of the requirements for the degree of Engineer.

Classification of the microstructural evolution stages of steam reform tubes with different chemical compositions using artificial intelligence – an NDT 4.0 approaching

Pedro Enrique Monforte Brandão Marques

November/2020

Advisor: Gabriela Ribeiro Pereira
Joint Advisor: Clara Johanna Pacheco

Course: Materials Engineering

Heat resistant austenitic stainless steels of HP grade are usually applied in severe operational conditions, suffering an aging process. This process, under abnormal operating conditions, can be accelerated, often generating the material failure by creep mechanisms. One of the applications for this material is in reforming furnace tubes. Previous studies have proposed methods of non-destructive inspection and classification of the microstructural evolution states of reform furnace tubes made of this material with specific chemical compositions. In this work, a new methodology for the classification of data from the inspection of tubes with varied chemical compositions is proposed, using techniques of artificial intelligence and applied statistics, based on the NDT 4.0 approach. The results of the use of the methodology proposed to train a unique model for tubes with two different chemical compositions showed that it was possible to reach an accuracy of 91.8% for the classification of data analyzed.

Keywords: Metallurgy, Austenitic Stainless Steels, Non-Destructive Testing, Artificial Intelligence, Data Analysis.

SUMÁRIO

1. INTRODUÇÃO 1
2. REVISÃO BIBLIOGRÁFICA 4
2.1. Fornos de reforma 4
2.2. Aços austeníticos HP 5
2.3. Envelhecimento dos aços HP modificados ao Nióbio 6
2.4. Ensaios não destrutivos magnéticos 8
2.5. END 4.0: A nova fronteira dos Ensaios Não Destrutivos 10
2.6. Aprendizado de máquina 14
2.7. Preparação dos dados 26
3. METODOLOGIA 34
3.1. Materiais analisados 34
3.2. Coleta de dados 35
3.3. Análise de compatibilidade de classes 36
3.4. Pré-processamento 37
3.5. Visualização dos dados 39
3.6. Treinamento e validação do modelo de classificação 40
4. RESULTADOS E DISCUSSÃO 42
4.1. Análise de compatibilidade de classes 42
4.2. Pré-processamento 42
4.2.1. Análise de normalidade 42
4.2.2. Análise de correlação 42
4.2.3. Análise de sobreposição 45
4.2.4. Análise de seleção de variáveis 46
4.3. Visualização dos dados 47
4.1. Treinamento dos modelos de classificação 49
4.1. Avaliação dos atributos 53
5. CONCLUSÃO 54
REFERÊNCIAS 55

1. INTRODUÇÃO

Nas últimas décadas, muitos debates vem ocorrendo na nossa sociedade acerca dos impactos socioambientais da queima de combustíveis fósseis como fontes energéticas. Em todo o mundo, inclusive no Brasil, isso tem levado a um vertiginoso aumento da demanda de fontes energéticas menos nocivas ecologicamente. Para obtê-las, novas tecnologias vem sendo desenvolvidas, como o hidrotratamento e hidrocraqueamento do óleo cru pesado, que baseiam-se em estabilizar derivados do petróleo com a utilização de hidrogênio, permitindo a produção de combustíveis com um controle superior de poluentes em sua combustão. Portanto, o aumento da demanda de novos combustíveis se reflete num aumento da demanda de hidrogênio como matéria-prima (ARENAS, 2017).

A obtenção do hidrogênio, sobretudo no Brasil, tem como peça fundamental os fornos de reforma a vapor. Isso ocorre, pois, utilizando hidrocarbonetos leves e vapor d´água, esses fornos tem a capacidade de produzir grandes quantidades de hidrogênio por meio da reação de reforma. Entre os materiais candidatos para a construção dos tubos dos fornos, o aço austenítico HP é largamente utilizado pela sua capacidade de resistir a essas severas condições dos fornos de reforma por longos períodos, com um custo mais competitivo que outros materiais com as mesmas propriedades. As colunas desse material são produzidas por meio de fundição por centrifugação, gerando assim uma estrutura composta por dendritas que crescem de forma transversal a parede dos tubos (BONACCORSI, 2014).

Apesar de suas excelentes propriedades termomecânicas, por estar continuamente sujeito a altas temperaturas e esforços, esse material sofre fenômenos de envelhecimento. Esse processo, sob condições operacionais anormais pode ser acelerado, podendo, nesses casos, gerar falhas por fluência, reduzindo a vida útil dos tubos. Esse envelhecimento microestrutural consiste na transformação dos carbetos primários interdendríticos e na precipitação secundária intradendrítica de carbetos e outras fases, gerando microestruturas características sequenciais já bem diferenciadas por trabalhos anteriores, recebendo a denominação de “estados de evolução microstrutural” (QUEIROZ, 2017). O problema descrito, associado ao processo de envelhecimento, portanto, é especialmente preocupante quando um estado de evolução microestrutural mais avançado é detectado em uma região do tubo onde se esperaria um estado inferior.

Mesmo as colunas estando sob condições de operação estáveis, devem ser feitas inspeções periódicas nas paradas de manutenção para avaliar o estado do material e estimar a vida residual dos tubos, o que atualmente é feito, em geral, via inspeção visual e medição de diâmetro ao longo do comprimento da coluna (ARENAS, 2017). Todavia, uma alternativa promissora para a caracterização microestrutural dos aços HP é empregar ensaios magnéticos não destrutivos que permitam associar as mudanças de magnetismo do material ao seu estado de evolução microestrutural, como demonstrado com sucesso por trabalhos recentes como (ARENAS, 2013).

Para gerar um método robusto e automático de classificação dos dados gerados pelos ensaios magnéticos em diferentes estados de tubos de reforma a vapor, algoritmos de aprendizado de máquina se mostram uma opção possível. Eles permitem criar modelos de classificação que atribuam a novos dados captados uma determinada classificação, chamada “classe”, a partir de um conjunto de dados de treinamento com classificações bem conhecidas, ação denominada predição. Esses modelos se baseiam em dados compostos por pontos estruturados em vetores, chamados de “exemplos” nos quais cada elemento é um parâmetro medido experimentalmente ou calculado, chamado de “atributo”.

Portanto, para realizar o desenvolvimento de modelos de aprendizado de máquina, é fundamental ter uma fonte confiável de dados experimentais para serem utilizados como dados de treinamento para o modelo. Essa fonte, no problema atual, deve vir de inspeções não destrutivas de tubos de reforma em operação em refinarias, ou amostras desses. Para se obter um modelo capaz de classificar todos os estados de evolução microestrututal dos tubos, portanto, são necessários dados correspondentes a um conjunto de tubos ou amostras de tubos que apresentem todos os estados em suas microestruturas. Em geral, isso é feito a partir de amostras retiradas de diferentes alturas dos tubos onde, pelas condições locais de temperatura e pressão, se esperaria ter um determinado estado, os quais são comprovados por ensaios de metalografia.

Trabalhos anteriores realizados no LNDC/UFRJ (ARENAS, 2017, DUTRA, 2019) desenvolveram metodologias de classificação baseados no ensaio de correntes parasitas com uma sonda híbrida, aplicando algoritmos de aprendizado de máquina para classificar os dados associados aos estados de evolução de tubos com uma mesma composição química. Todavia, observa-se que existe uma variação considerável entre os dados para cada estado de evolução microestrutural das inspeções de amostras com composições químicas diferentes. Isso é esperado, uma vez que as propriedades físicas do material, em especial as elétricas e magnéticas, tem grande dependência da sua composição química e das variações microestruturais consequentes dessa variação composicional, ainda que o método de conformação seja

o mesmo. Essa variedade de propriedades gera, então, uma oscilação das respostas magnéticas dos mesmos estados microestruturais de materiais com diferentes composições. Esse fato é extremamente prejudicial no momento em que se deseja generalizar modelos de classificação automáticos que tem um bom desempenho para materiais com uma certa composição.

No trabalho atual, é proposta uma nova metodologia para gerar modelos de classificação dos mesmos estados de evolução microestrutural para tubos com diferentes composições, um estágio fundamental para a generalização do método de inspeção. Para isso, foi treinado um modelo único com dados provenientes de amostras de duas composições químicas distintas. Esse modelo deverá ter um desempenho de classificação superior a dois modelos desenvolvidos para cada composição individualmente, e obter métricas gerais de avaliação de desempenho aceitáveis para aplicação em casos reais da indústria.

2. REVISÃO BIBLIOGRÁFICA

2.1. Fornos de reforma

A obtenção de hidrogênio pela indústria petroquímica é realizada em fornos de reforma. A produção do hidrogênio ocorre mediante reações endotérmicas entre vapor de água e hidrocarbonetos (majoritariamente o metano) em processos catalíticos que envolvem grandes quantidades de calor. A reação de craqueamento ocorre no interior dos tubos de reforma por onde o gás de reação escoa, em temperaturas elevadas, através de um leito catalisador (SILVEIRA e MAY, 2006).

As colunas de reforma são submetidas a temperaturas de parede que variam entre 600°C e 1.000°C com uma pressão interna entre 1,0 e 3,5 MPa. O comprimento destas colunas varia de 10 a 14 m, o diâmetro externo entre 100 e 200 mm e a espessura de 10 a 20 mm (SILVEIRA e MAY, 2006).

As altas pressões e temperaturas às quais as colunas dos fornos de reforma estão submetidas as tornam componentes extremamente críticos e, sendo assim, é inevitável o aparecimento de danos em serviço, podendo levar a falhas catastróficas em certas condições de operação. Devido a essas condições de trabalho, os aços austeníticos de grau HP são os mais indicados para a fabricação dos tubos de reforma, por apresentarem propriedades como alta resistência ao calor, fluência e corrosão a custos competitivos (BONACCORSI, 2014).

O arranjo e as características operacionais de um forno de reforma variam conforme a tecnologia proprietária do projetista e a descarga mássica requerida para o gás de processo. As amostras estudadas por essa dissertação pertencem a um forno *top-fired*, no qual os queimadores são posicionados no teto da câmara de radiação e a disposição das colunas é vertical (Figura 1.a). Nesta figura também se apresenta um exemplo de perfil de temperatura (Figura 1.b).

A vida útil nominal dos tubos é de aproximadamente 100.000 horas e, durante esse tempo, decorrentes do processo de envelhecimento ao qual eles são submetidos, ocorrem mudanças microestruturais no aço, consequentemente alterando suas propriedades mecânicas. Falhas nos tubos podem causar paradas não planejas e até perda na produção de hidrogênio podendo ser extremamente custosas.

Uma das principais causas da falha dos tubos usados em fornos de reforma é causada por fluência (QUEIROZ, 2012). Portanto é indispensável que haja uma programação de substituição preventiva, visando prevenir quaisquer falhas sérias e

buscando também o maior aproveitamento possível do tempo em campanha dos tubos (SWAMINATHAN *et al.*, 2008).

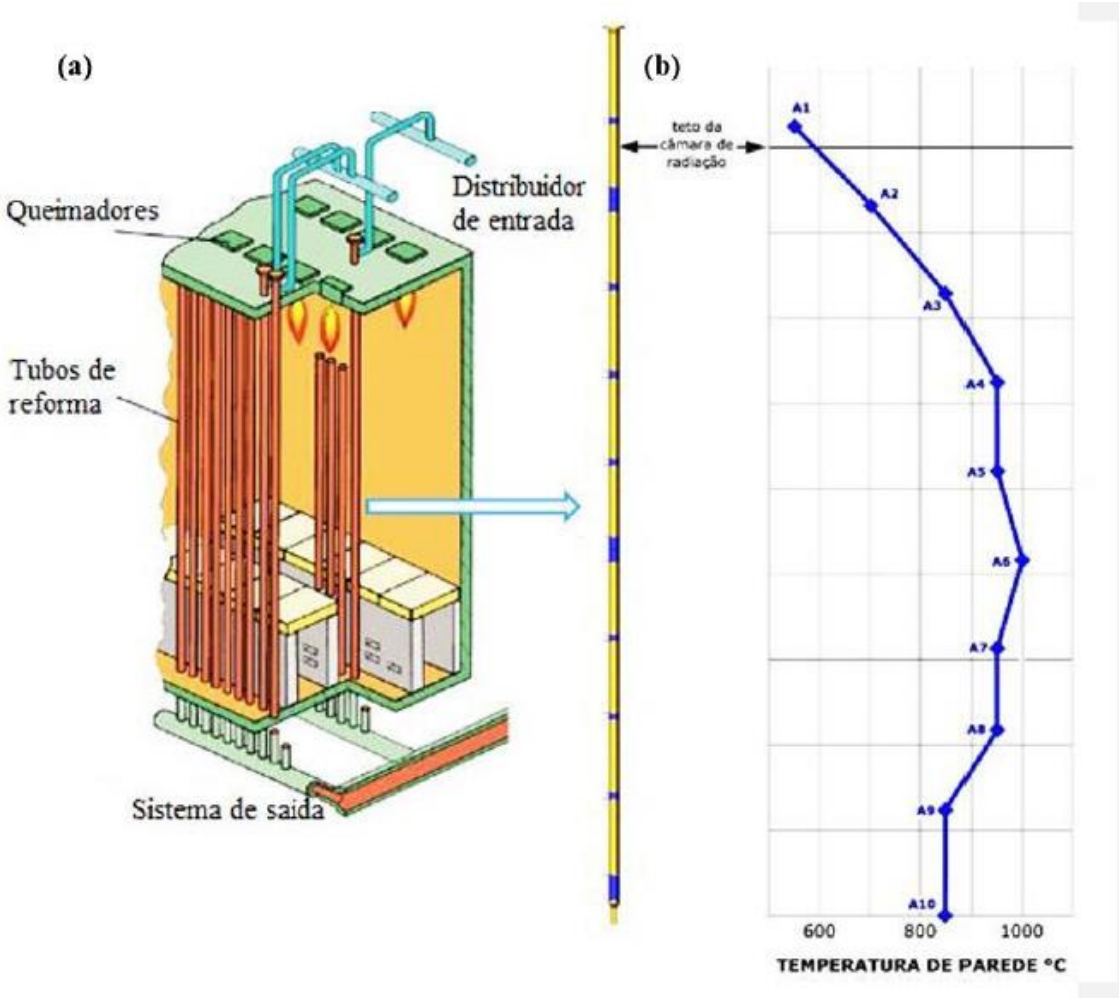


Figura 1. (a) Esquema de um forno de reforma top-fired. (b) Exemplo da distribuição de temperatura em um tubo de reforma. Adaptado de (ARENAS, 2013).

2.2. Aços austeníticos HP

Grande parte dos aços utilizados na indústria petroquímica são ligas austeníticas compostas de ferro, cromo e níquel. As ligas da classe Fe-Cr-Ni possuem alta resistência à fluência, suportam grandes carregamentos, aceitam variações de temperatura moderadas e podem ser utilizados em condições tanto oxidantes como redutoras. Como o níquel atua como estabilizador da austenita, as ligas da classe Fe-Ni-Cr com maiores teores de níquel possuem microestruturas mais estáveis, tendo

menor probabilidade de formação de fases deletérias, resultantes da exposição do aço a temperaturas altas (TOMISKA, 2004).

Os aços inoxidáveis austeníticos resistentes ao calor utilizados nos fornos de reforma a vapor vem aprimorando suas propriedades mecânicas através de modificações na composição química. Isso é necessário, pois, devido às severas condições de serviço, os tubos tornam-se susceptíveis a falhas por fluência, a descarbonetação, a corrosão e a fadiga térmica.

A nomenclatura dessa classe de aços fornece as informações sobre qual temperatura é indicado para uso, representado pela primeira letra, e qual a relação dos elementos níquel-cromo, representada pela segunda letra. A primeira letra pode ser C para aços aplicados em ambientes aquosos corrosivos com temperatura de uso até 650°C e H para aços resistentes a altas temperaturas com uso acima de 650ºC. A segunda letra varia de A a Z, onde A representa a menor quantidade de Ni e Z representa a maior quantidade de Ni. Por exemplo, o aço HP é indicado para usos a altas temperaturas e possui de 24-28% de Cr e 33-37% de Ni (ARENAS, 2013).

As propriedades mecânicas dos aços HP podem ser melhoradas através de modificações das suas microestruturas, com a adição de elementos químicos como nióbio, titânio, zircônio, ítrio e tungstênio (ASTM, 2006).

2.3. Envelhecimento dos aços HP modificados ao Nióbio

Aços inoxidáveis austeníticos passam por uma evolução microestrutural quando estão em serviço, sendo esta ditada pelo tempo de exposição e pela temperatura de trabalho (BARBABELA, 1990).

Na Tabela 1 serão descritas as principais características de cada estado de evolução microestrutural de um aço HP modificado ao Nb quando submetido a diferentes temperaturas de operação após vários milhares de horas em serviço (Queiroz, 2017). Microestruturas dos estados podem ser vistos na Figura 2.

Tabela 1. Estados de evolução do aço HP. Adaptado de (Queiroz, 2017).

Estado	Microestrutura	Temperatura (ºC)
I	Matriz austenítica com precipitação primária de carbetos metálicos $M_{23}C_6$ ou M_7C_3, com predominância de Cr.	T < 600
II	Intensa precipitação secundária de carbetos de cromo, $M_{23}C_6$, que podem precipitar em contornos de grão ou de inclusões.	600 < T < 700
III	Carbetos primários transformam completamente sua morfologia eutética em blocos compactos, além de finos carbetos secundários serem dispersos na matriz.	700 < T < 800
IV	Matriz austenítica com uma nuvem de precipitados finos.	800 < T < 900
V	A quantidade de precipitados secundários no interior da matriz diminui e o tamanho médio dos precipitados aumenta substancialmente.	T ~ 900
VI	Evolução da condição descrita anteriormente, sendo mais intenso o coalescimento de carbetos secundários e o aparecimento de microvazios.	T > 900

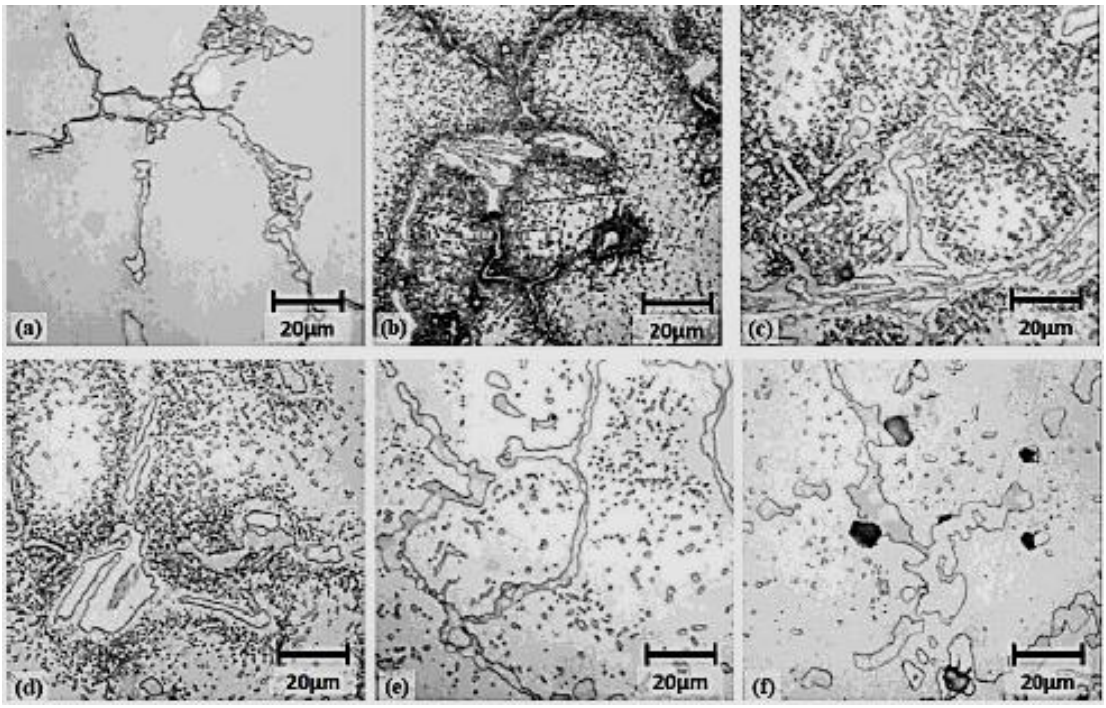


Figura 2. Estados de evolução microestrutural de um aço austenítico HP-Nb. Estados: (a) I, (b) II, (c) III, (d) IV, (e) V e (f) VI. Adaptada de (ARENAS, 2017).

Como demonstrado por trabalhos anteriores (ARENAS, 2013, MARTINS, 2000), as condições microestruturais e a presença de defeitos em aços austeníticos resistentes ao calor alteram a resposta magnética dessa classe de aços. Por conta disso, os dados correspondentes a essa resposta podem ser obtidos através de ensaios não destrutivos (END) magnéticos.

2.4. Ensaios não destrutivos magnéticos

O eletromagnetismo, compreendido como o conjunto de efeitos físicos resultantes da interação da eletricidade e do magnetismo, formam a base de um grande número de métodos de END. Todos estes métodos são indiretos; uma propriedade do material é medida como uma variação de uma propriedade elétrica ou magnética, como condutividade ou resistência elétrica. Os métodos são aplicáveis a todos os materiais condutores de eletricidade e incluem diferentes tipos de ensaio como: testes de corrente parasita, inspeção de partículas magnéticas, vazamento de fluxo magnético, entre outros. Os testes de indução eletromagnética são aplicados a todos os estágios de fabricação e processamento de metais e ligas.

As técnicas baseadas em fenômenos eletromagnéticos têm certas vantagens sobre outras convencionais, como o ultrassom, tornando-as mais adequadas para muitas aplicações. Entre estas vantagens está o fato de não ser necessário nenhum meio de acoplamento e de não necessitarem sequer de estar em contato com a superfície inspecionada, o que permite o uso de velocidades de inspeção elevadas. Além disso, estas técnicas permitem que os defeitos sejam localizados e dimensionados com altos níveis de repetibilidade e com a possibilidade de realizar análises em tempo real. Ademais, o uso de sistemas computadorizados para aquisição e análise de dados torna a interpretação dos resultados relativamente simples (NDT.NET, 2020).

Um dos principais ensaios não destrutivos magnéticos é o ensaio de correntes parasitas (ou “*eddy current test*”, ou ECT). O ensaio de corrente parasita é uma técnica de inspeção não destrutiva baseada na indução de uma corrente elétrica alternada no material a ser inspecionado e na observação da interação entre o campo magnético gerado pela sonda e o campo magnético gerado pela corrente induzida no material inspecionado (NDT.ED, 2020).

O princípio básico do funcionamento por corrente parasita é semelhante ao de um transformador, onde a bobina da sonda corresponde ao componente primário e o material inspecionado ao componente secundário do transformador. De acordo com a lei de Ampère, quando uma bobina é excitada por uma corrente alternada, é gerado um campo magnético alternado primário. Ao aproximar esta bobina da superfície de um material condutor, o campo gerado pela bobina induz uma corrente elétrica no material, denominada "corrente parasita". Estas correntes geradas dão origem a um novo campo magnético, conhecido como campo secundário, que se opõe às variações do campo primário. A interação entre os campos primário e secundário modifica a impedância elétrica da bobina, cuja variação é a variável física monitorada durante o teste (CAMERINI, 2018), como pode ser visto na Figura 3 (NDT.ED).

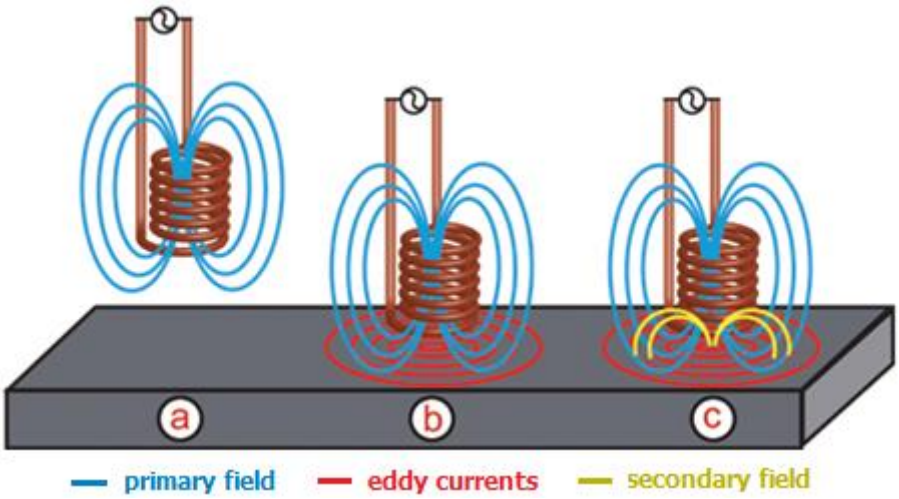


Figura 3: Princípio do ensaio de correntes parasitas. Adaptado de (NDT.ED, 2020).

Se uma região do material inspecionado apresentar alguma descontinuidade, essa região apresentará um fluxo de correntes parasitas diferente que as regiões ao redor, pois a região com a descontinuidade apresenta uma resistência diferente à passagem de correntes elétricas. Esta diferença no fluxo de correntes parasitas resulta em uma mudança na impedância da bobina, que é constantemente monitorada, sendo possível diferenciar entre regiões com e sem descontinuidade. Essas descontinuidades podem ser vazios, defeitos, inclusões, entre outras. Por outro lado, variações microestruturais do material também geram alterações no fluxo de correntes parasitas, modificando localmente a permeabilidade e condutividade elétrica, tornando possível a comparação da microestrutura entre diferentes materiais ou regiões de materiais com microestruturas ou dispersões de fase complexas. O sinal medido dependerá das características geométricas da bobina, da frequência de ensaio, das propriedades

elétricas e magnéticas do material, das dimensões da amostra e da existência de descontinuidades na amostra (NDT.ED).

No teste das correntes parasitas, a utilização de padrões de referência no ajuste e calibração do equipamento é essencial, já que os sinais são afetados por muitas variáveis diferentes. Em consequência, pequenas alterações na configuração do equipamento podem alterar drasticamente a aparência de um sinal, definindo esta técnica como comparativa. Como na maioria dos métodos de ensaios não destrutivos, a informação mais útil é obtida quando se comparam os resultados de um objeto desconhecido com os resultados de um objeto similar, com características e defeitos bem conhecidos (ARENAS, 2017).

Nos últimos anos, os ensaios não destrutivos magnéticos tem recebido muita atenção devido às suas grandes potencialidades para aplicações cada vez mais complexas e avançadas. Uma delas é o desenvolvimento de sistemas de inspeção automáticos, acoplados com sistemas de classificação e transmissão de dados em alta velocidade. Isso coloca esse tipo de técnica no centro da nova fronteira dos END nesse século, integrada com as recentes inovações produtivas com foco na automação, tendência que recebeu o nome de END 4.0.

2.5. END 4.0: A nova fronteira dos Ensaios Não Destrutivos

Historicamente, a revolução industrial pela qual a humanidade passou é dividida em três fases: (1) a invenção da máquina a vapor (mecanização), (2) a eletricidade (produção em massa) e (3) a revolução microeletrônica (automação), como são ilustradas na Figura 4. No campo tecnológico dos ensaios não destrutivos, um desenvolvimento semelhante e equivalente aconteceu: ferramentas como lentes ou estetoscópios permitiram aguçar os sentidos humanos, a conversão das ondas tornou visível o invisível e assim ofereceu um "olhar" sobre os componentes e finalmente foram cruciais para a automação, digitalização e reconstrução. Durante todo o desenvolvimento industrial, os END foram decisivamente responsáveis pela qualidade e, portanto, pelo sucesso dos produtos fabricados pelo ser humano (VRANA, 2019).

A indústria está agora a falar de uma quarta revolução: a fase da informatização, da digitalização e do trabalho em rede da produção industrial, gerando os chamados sistemas “físico-cibernéticos”. A chamada "Indústria 4.0" opera com base em alguns princípios fundamentais: interoperabilidade, transparência da informação, assistência técnica e decisões descentralizadas.

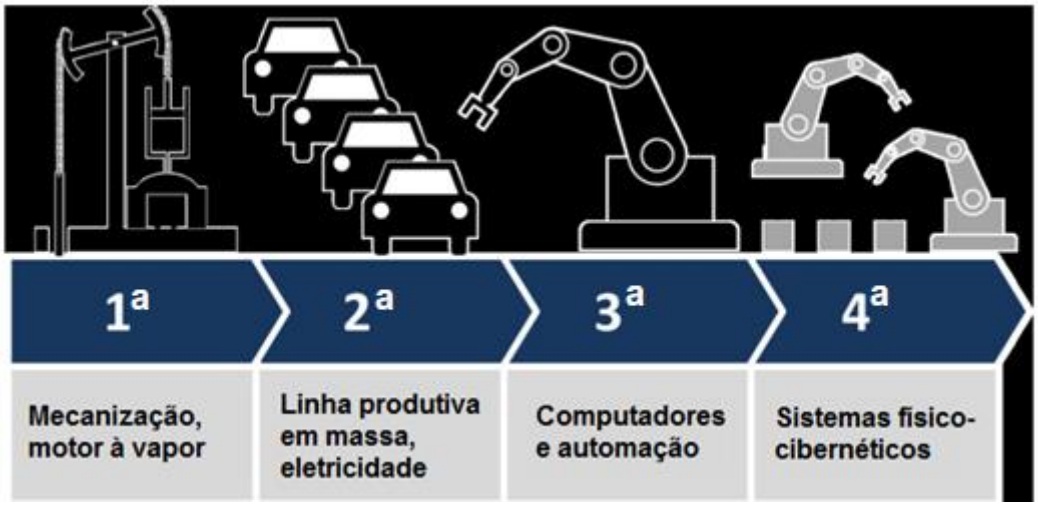


Figura 4: Fases da revolução industrial. Adaptada de *Creative Commons*.

As máquinas devem ser capazes de se comunicar umas com as outras, principalmente através da Internet das Coisas (IoT). Isto deve permitir-lhes criar cópias virtuais do mundo físico através de dados de sensores, ajudar os humanos fornecendo uma visualização agregada de informações para uma melhor tomada de decisões e garantir que os sistemas sejam capazes de encontrar soluções de forma independente e implementar as funções para as quais foram concebidos (VRANA, 2019).

Como nas outras revoluções anteriores, os métodos de END serão fundamentais para o sucesso desta quarta revolução, fornecendo a base de dados necessária para a realimentação em um ambiente de produção em rede. Um dos aspectos mais essenciais do então chamado END 4.0 é, portanto, a possibilidade de avaliação e decisão automatizada. Para os END, isto levará a grandes mudanças. A disponibilidade de IoT, sistemas cibernéticos, formação de nuvens, plataformas digitais e inteligência artificial, incluindo redes neurais, oferecem grandes possibilidades e desafios em todas as áreas de produção, serviços e sociedade (LINK e RAINER, 2018).

O END 4.0 compreende todos os aspectos dentro da Indústria 4.0 para a inspeção do material desde o bruto até o processado, inspecionado e classificado até a entrega ao cliente e durante toda sua vida operacional. Informações relevantes ao longo de todo o processo de produção, a garantia de qualidade aos clientes pode, assim, ser fornecida e, principalmente, cobrada. O material desde o início do processo pode estar equipado com um sensor, que guarda toda a informação sobre o seu futuro, escolhendo o seu caminho através da fabricação e inspeção. As tendências impulsionadas pelos dados da Indústria 4.0 são alimentadas pela informação fornecida pelos sensores. Estes sensores são omnipresentes e vão desde pontos de dados discretos como temperatura, tempo, posição, etc. até aos sensores mais complexos que fornecem uma profundidade de informação muito maior. Um exemplo desses sensores encontrados no END são

sistemas de Tomografia Computadorizada (CT) - provavelmente o sensor que tem o potencial de fornecer o mais alto nível de informação sobre o produto e processos de uma empresa (SCHULENBURG, 2019).

Para resumir as potencialidades do END 4.0 de forma esquemática, uma análise baseada na matriz FOFA (do inglês, "*SWOT*") foi realizada por (LINK e RAINER, 2018), e pode ser observada de forma adaptada na Tabela 2.

Entre os avanços tecnológicos associados à Indústria 4.0 que mais vem encontrando aplicações integradas a sistemas de inspeção não destrutiva são os algoritmos de aprendizado de máquina. A partir deles, os programas computacionais podem analisar milhões de pontos de dados e usar software de reconhecimento automático de defeitos para não só localizar defeitos ou variações microestruturais em uma peça, mas também reconhecer o tipo de descontinuidade apresentada automaticamente e em tempo real. Esta informação é crucial, pois pode permitir aos fabricantes mudar seu processo de fabricação ou indicar manutenções, se necessário, garantindo a flexibilidade que pode economizar dinheiro das empresas em longo prazo. Em última análise, novas tecnologias como estas têm um enorme impacto no END hoje e continuam a avançar e crescer à medida que mais empresas estão a descobrir como utilizá-lo da forma mais eficiente possível.

No âmbito desse trabalho, vimos que as condições microestruturais e a presença de defeitos em aços austeníticos resistentes ao calor alteram a resposta magnética dessa classe de aços, a qual pode ser lida e transformada em dados experimentais através do uso de ensaios não destrutivos magnéticos. Esses dados, uma vez digitalizados, entram na lógica do END 4.0, podendo ser utilizados para o desenvolvimento de sistemas de classificação automáticos para os diferentes estados de evolução microestrutural do aço HP, baseados em aprendizado de máquina (ARENAS, 2013, DUTRA, 2019). Para continuar a dissertação sobre esses sistemas, que são o objeto desse trabalho, é necessário compreender os princípios e funcionamento da metodologia de aprendizado de máquina.

Tabela 2: Matriz FOFA do END 4.0. Adaptada de (LINK e RAINER, 2018).

END 4.0	Fatores positivos	Fatores negativos
Fatores internos	• Individualização dos produtos; • Nuvens e plataformas podem ser aplicadas a transmissão de dados experimentais; • Integração de todas as etapas de processamento em operação; **FORÇAS**	• Alta qualificação profissional necessária. • Custo de investimento em equipamentos e qualificação; • Pouco suporte para empresas de médio porte. **FRAQUEZAS**
Fatores externos	• Aumento de produção; • Conexão digital entre todas as operações internas e externas; • Sistemas de inspeção de qualidade automatizados; • Análise de dados em tempo real incluindo *Big Data*; **OPORTUNIDADES**	• Aumento do desemprego; • Potencial perda de conhecimento estratégico para competidores via internet; • Falta de capital humano qualificado; **AMEAÇAS**

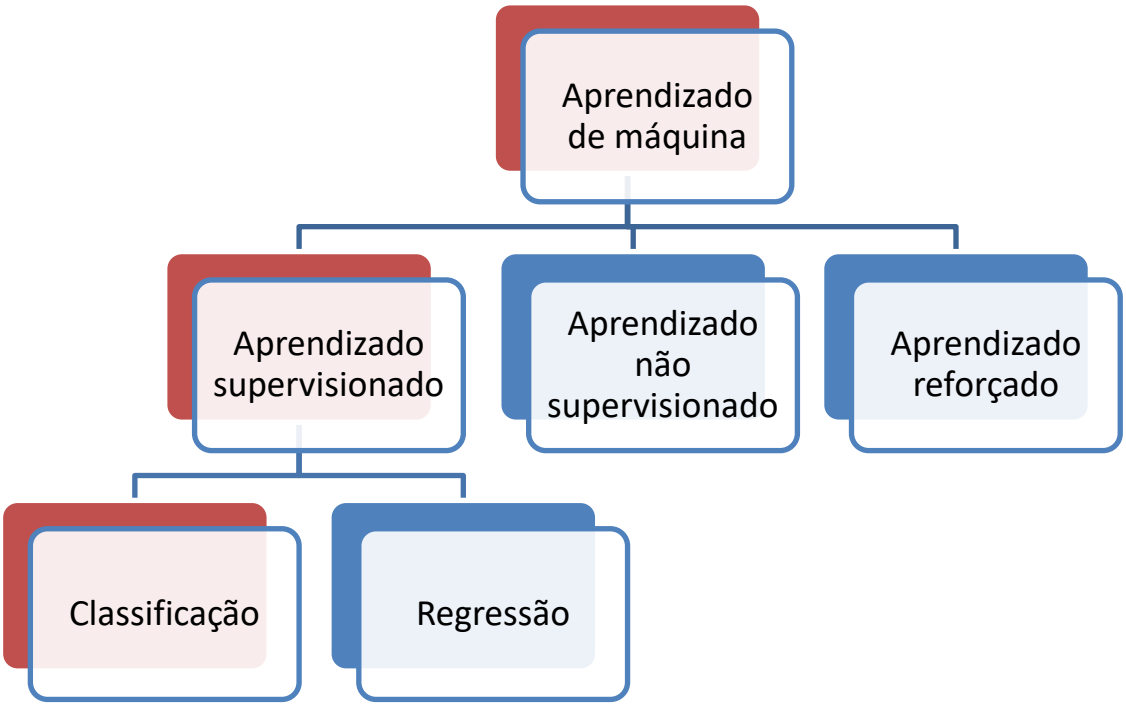
2.6. Aprendizado de máquina

2.6.1. Definições

Aprendizado de Máquina (AM) é uma área de Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores (REZENDE, 2003).

O Aprendizado de Máquina pode ser classificado em três categorias, mas, no entanto, com base na situação, estas categorias podem ser combinadas para alcançar os resultados desejados para determinadas aplicações (DANGETI, 2017). Essas categorias podem ser observadas na Tabela 3.

Tabela 3: Categorias de Aprendizado de Máquina.



- O aprendizado supervisionado inclui um conjunto de métodos capazes de treinar máquinas para aprender a relação entre outras variáveis e uma variável alvo, semelhante à forma como um professor fornece avaliações aos alunos sobre o seu desempenho. Em outras palavras, treinar a máquina para rotular novos dados a partir de um conjunto de dados previamente rotulado. Os principais

segmentos dentro da aprendizagem supervisionada são os modelos de classificação e de regressão (DANGETI, 2017). Nesse trabalho, daremos foco ao uso de modelos de classificação supervisionada.

- No aprendizado não supervisionado, os algoritmos aprendem por si próprios sem qualquer supervisão ou sem qualquer variável alvo fornecida. É uma questão de encontrar padrões e relações ocultas nos dados fornecidos (DANGETI, 2017).
- O aprendizado reforçado, por outro lado, permite que a máquina ou agente aprenda seu comportamento com base na resposta do ambiente. A máquina toma uma série de ações decisivas sem supervisão e, no final, uma recompensa será dada, ou +1 ou -1. Com base na recompensa final, o agente reavalia os seus caminhos. Os problemas de aprendizagem de reforço estão mais próximos da metodologia de IA do que os algoritmos de aprendizagem de máquina frequentemente utilizados (DANGETI, 2017).

Portanto, em todos os casos descritos acima o objetivo do aprendizado de máquina é treinar algoritmos de classificação (chamados de “classificadores”), capazes de classificar dados novos com o objetivo de predizer corretamente os rótulos de cada um. Após isso, o classificador pode ser avaliado considerando sua precisão, compreensibilidade ou grau de interesse, velocidade de aprendizado, requisitos de armazenamento, grau de compactação ou qualquer outra propriedade desejável que determine quão bom e apropriado ele é para a tarefa em questão (REZENDE, 2003).

Formalmente, os dados passíveis de serem classificados via aprendizado de máquina são nomeados de exemplos (do inglês, “*instances*”). Um exemplo descreve um dado objeto de interesse, tal como um determinado paciente, exemplos médicos sobre uma determinada doença ou histórico de clientes de uma dada companhia. Todo exemplo é definido por um conjunto fixo de atributos (comumente chamadas em inglês de “*features*”), os quais são variáveis aleatórias que determinam propriedades individuais mensuráveis de determinado exemplo. Se os exemplos de determinado problema de AM fossem pacientes de um hospital, seus atributos poderiam ser sua temperatura corporal, peso, idade, entre outros (REZENDE, 2003).

Portanto, em outras palavras, o objetivo dos classificadores é predizer rótulos corretamente para cada exemplo, considerando o conjunto de atributos que o definem. Esses rótulos são comumente chamados de classes. No caso de problemas de classificação, as classes são pertencentes a um conjunto discreto (nominal) $\{C_1, C_2, ..., C_k\}$, ou, no caso de problemas de regressão, são pertencentes a conjuntos contínuos, em geral dos números reais. Na analogia dos exemplos serem pacientes de um hospital, uma classe discreta poderia ser qual doença o paciente tem dentro de um conjunto finito

de doenças possíveis, e uma classe contínua poderia ser a probabilidade do paciente ter uma determinada doença (REZENDE, 2003).

É importante ressaltar que, em problemas de aprendizado supervisionado, como é necessário receber um conjunto de exemplos previamente classificados para treinar o classificador, a classe de cada exemplo também pode ser considerada um atributo desse exemplo (REZENDE, 2003), como mostra a Figura 5.

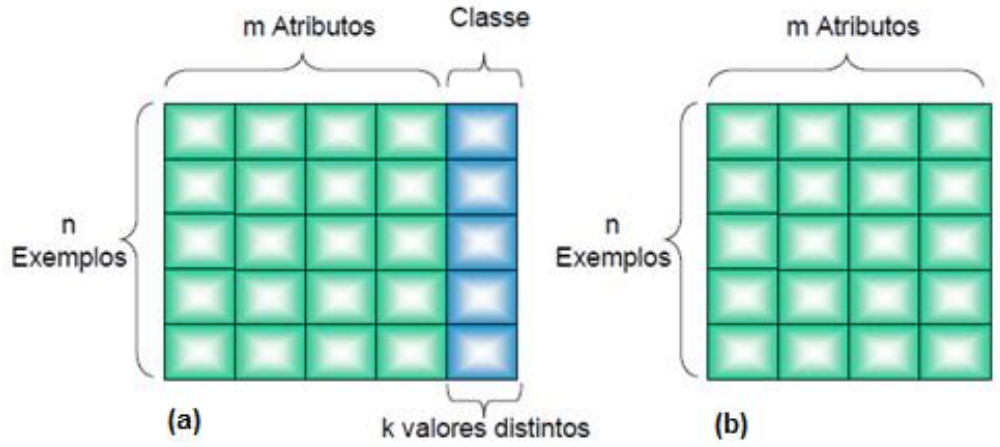


Figura 5: (a) Conjunto de exemplos para um problema de aprendizado supervisionado; (b) Conjunto de exemplos para um problema de aprendizado não supervisionado. Adaptado de (REZENDE, 2003).

A partir dos conceitos apresentados, é possível então interpretar cada exemplo como um vetor, cujas m coordenadas são seus atributos. Assim, pode ser definido um espaço vetorial m-dimensional denominado espaço de descrição, onde os vetores de cada exemplo estão contidos. Esse conceito é importante, pois, em problemas de classificação supervisionada, cada ponto no espaço de descrição pode ser rotulado com a classe associada, o que permite o emprego de diversas ferramentas matemáticas como, por exemplo, da álgebra linear (REZENDE, 2003).

2.6.2.Classificadores

Como visto anteriormente, dado um conjunto de treinamento, o algoritmo responsável pelo aprendizado de máquina gera como saída um classificador (também chamado de “hipótese” ou “descrição de conceito”) de forma que, dado um novo exemplo, ele possa predizer precisamente sua classe (REZENDE, 2003).

Ou seja, cada exemplo é um par $(x, f(x))$, onde x é a entrada, $f(x)$ é a saída (sendo f uma função a princípio desconhecida). Como também visto na sessão anterior,

se $y = f(x)$ assume valores discretos $y \in \{C1, C2, \ldots, Ck\}$, trata-se de um problema de classificação, e se $y = f(x)$ assume valores reais, trata-se de um problema de regressão. O processo de aprendizado, também chamado de indução ou inferência indutiva trata de, em outras palavras, dada uma coleção de exemplos de $f(\bullet)$, retornar uma função $h(\bullet)$ que aproxima $f(\bullet)$, ou seja, $h(x) \cong f(x)$. Dessa forma, $h(\bullet)$ é denominada uma hipótese sobre a função objetivo $f(\bullet)$ (REZENDE, 2003).

Em geral, um conjunto de exemplos é dividido em dois subconjuntos disjuntos: o conjunto de treinamento, que é usado para o aprendizado do classificador; e o conjunto de teste, que é usado para medir o grau de efetividade do aprendizado. Os subconjuntos são disjuntos para assegurar que as medidas obtidas utilizando o conjunto de teste sejam de um conjunto diferente do usado para realizar o aprendizado, tornando a medida estatisticamente válida (REZENDE, 2003). Na prática, os exemplos geralmente são divididos aleatoriamente em 70%-30% ou 80%-20% nos conjuntos de treinamento e teste, respectivamente, como pode ser visto na Figura 6. Assim, os dados de treinamento são utilizados para construir o modelo e os dados de teste para verificar sua eficácia (DANGETI, 2017).

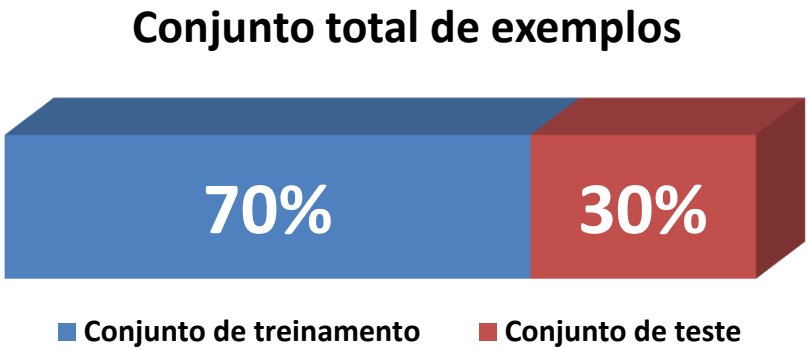


Figura 6: Exemplo de partição do conjunto total de exemplos em conjunto de treinamento e conjunto de testes.

Após treinar um classificador, é possível validá-lo no conjunto de treinamento bem como no conjunto de teste. É usual denominar as medidas de desempenho de um classificador efetuadas sobre o conjunto de treinamento como aparentes (também conhecidas como medidas de re-substituição) e as medidas efetuadas sobre o conjunto de teste como medidas reais (ou verdadeiras). Por exemplo, caso a medida seja o erro, usa-se a nomenclatura de erro aparente e o erro verdadeiro. Para a maioria das hipóteses, a medida aparente é um estimador ruim do seu desempenho futuro, uma vez

que ela tem a tendência de possuir um viés otimista. Portanto, em geral, o erro aparente é menor que o erro verdadeiro (REZENDE, 2003).

Cada modelo tem componentes de erro de viés e de variância, além de ruído branco. Os erros da componente de viés provêm de suposições errôneas no algoritmo de aprendizagem. Um algoritmo com elevado viés pode fazer com que um classificador não seja capaz de reproduzir as relações relevantes entre os atributos e as classes alvo dos exemplos. Por outro lado, os erros da componente de variância vêm de uma sensibilidade excessivamente alta do classificador a pequenas flutuações no conjunto de treinamento, ou seja, está relacionado com a incapacidade do algoritmo de reproduzir a variação entre diferentes conjuntos de exemplos usados como conjuntos de treinamento. Uma ilustração das componentes de erro de viés e variância pode ser vista na Figura 7 (DANGETI, 2017).

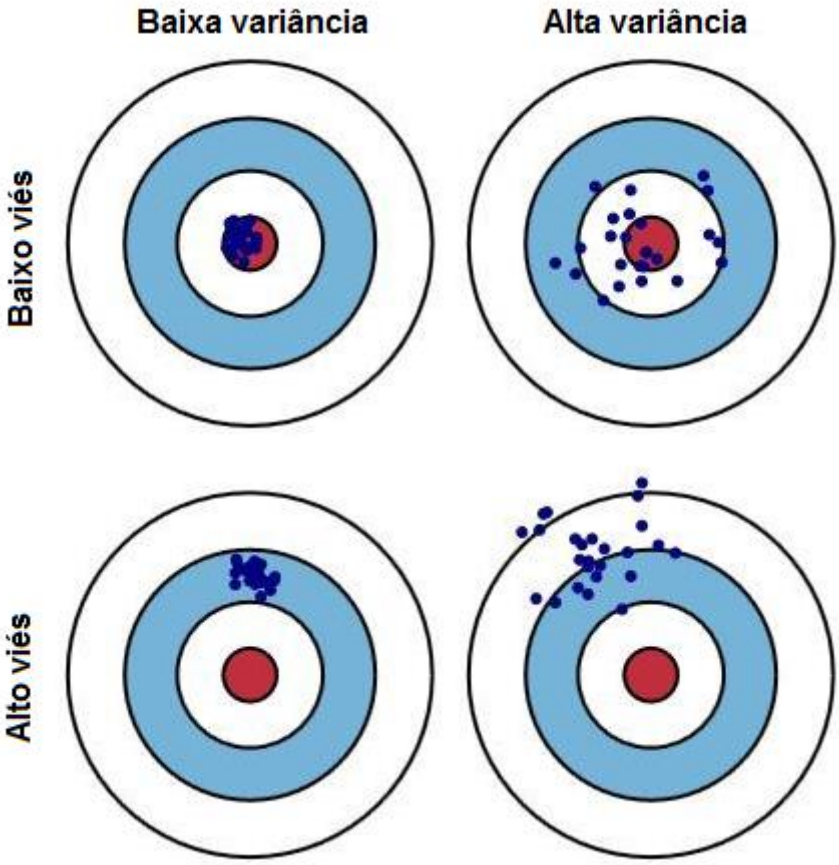


Figura 7: Ilustração das componentes de erro de viés e variância.

Buscando reduzir o erro, é possível utilizar um modelo complexo, ou seja, que é capaz de reduzir consideravelmente o erro de previsão no conjunto de treinamento, mas, ao mesmo tempo, que não é tão generalizável a ponto de apresentar um bom

resultado no conjunto de teste. Nesse caso, opta-se por estimar um modelo com viés baixo e variância alta. Ou seja, no conjunto de treinamento, o modelo irá ter um erro de previsão baixo, mas quando avaliamos o modelo em um novo conjunto de dados (conjunto de teste), o erro sobe consideravelmente, e por isso dizemos que sua variância é alta. Este é o típico caso de sobreajuste (do inglês, "*overfitting*"). Por outro lado, é possível utilizar um modelo simples, ou seja, que é bem generalizável a novos dados, mas não reduz consideravelmente o erro de previsão no conjunto de treinamento. Nesse caso, opta-se por um modelo com viés mais alto, mas variância baixa. Assim, iremos permitir um erro de previsão maior no conjunto de treinamento, de modo a obter um erro de previsão semelhante (com pouca variância) no conjunto de teste. Esse é o típico caso de sub-ajuste (do inglês, "*underfitting*"). A diferença entre esses dois casos pode ser ilustrado na Figura 8 (DANGETI, 2017).

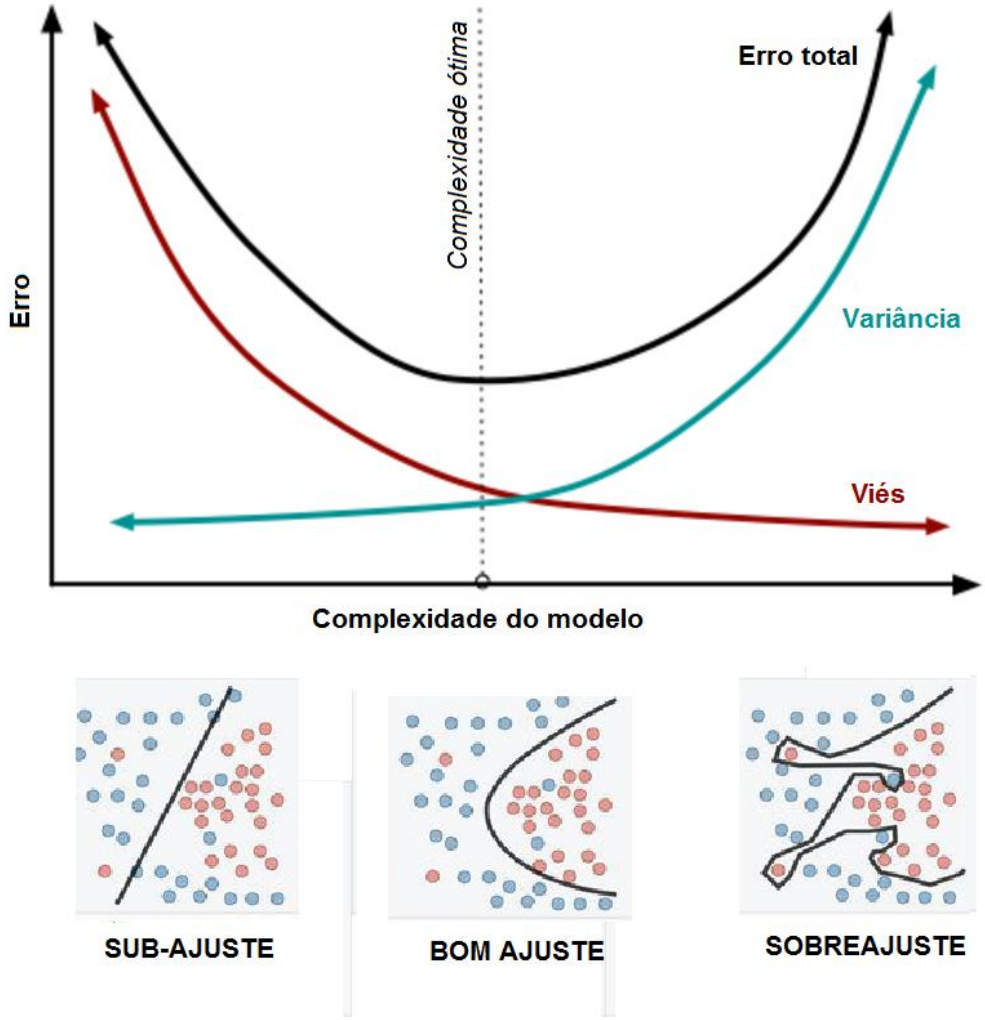


Figura 8: Ilustração da relação entre o erro e a complexidade dos modelos de aprendizado de máquina.

Além disso, para evitar erros de viés por utilizar o conjunto de treinamento durante o processo de validações dos classificadores, alguns autores como (DANGETI, 2017) apresentam outra metodologia de partição do conjunto de exemplos. Essa envolve adicionar um conjunto de validação igualmente disjunto ao conjunto de treinamento e teste, como mostra a Figura 9, para gerar uma validação não enviesada do modelo antes da realização do teste em si.

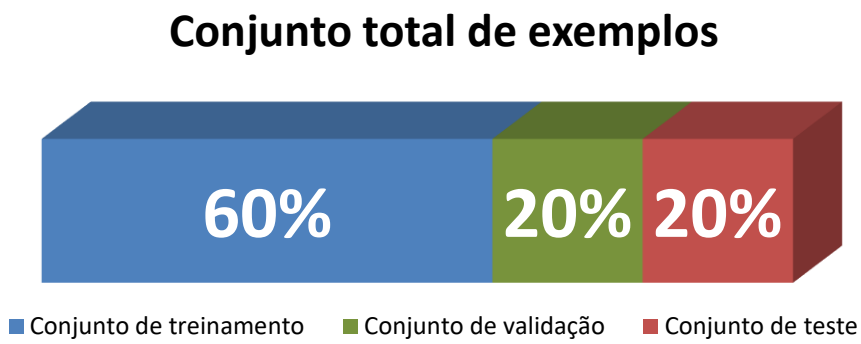


Figura 9: Exemplo de partição do conjunto total de exemplos em conjunto de treinamento, conjunto de validação e conjunto de testes.

2.6.3. Validação cruzada

Uma técnica alternativa para avaliar a capacidade de generalização dos modelos treinados para dados novos, isto é, para medir a probabilidade do modelo classificar corretamente novos dados medidos é a validação cruzada (YADAV, 2016), também conhecido como *k-fold cross validation* (do inglês).

O objetivo dessa técnica é definir um conjunto de dados para testar o modelo durante a fase de treinamento. Ele envolve a partição do conjunto de dados de treinamento em k subconjuntos, sendo k um número natural, onde um subconjunto é usado como conjunto de validação para testar o desempenho do modelo gerado usando os outros k-1 subconjuntos como conjunto de treinamento, como pode ser observado na Figura 10. Esse processo é repetido k vezes, obtendo-se um valor de acurácia por iteração. Ao fim, uma medida final de acurácia do modelo é obtida pela média aritmética da acurácia de cada iteração. Esse método de validação é muito interessante pois, como cada exemplo presente nos dados aparece em um conjunto de validação apenas uma vez e nos conjuntos de treinamento k-1 vezes, a acurácia e eficiência da validação são aumentadas, enquanto o viés estatístico e a variância são reduzidos (DANGETI, 2017). O valor ótimo para o parâmetro k depende de muitos fatores, como o tamanho do

conjunto de treinamento, e muitas vezes é escolhido subjetivamente. Todavia, o valor de k = 10 é um valor usualmente recomendado pela bibliografia como padrão (DANGETI, 2017).

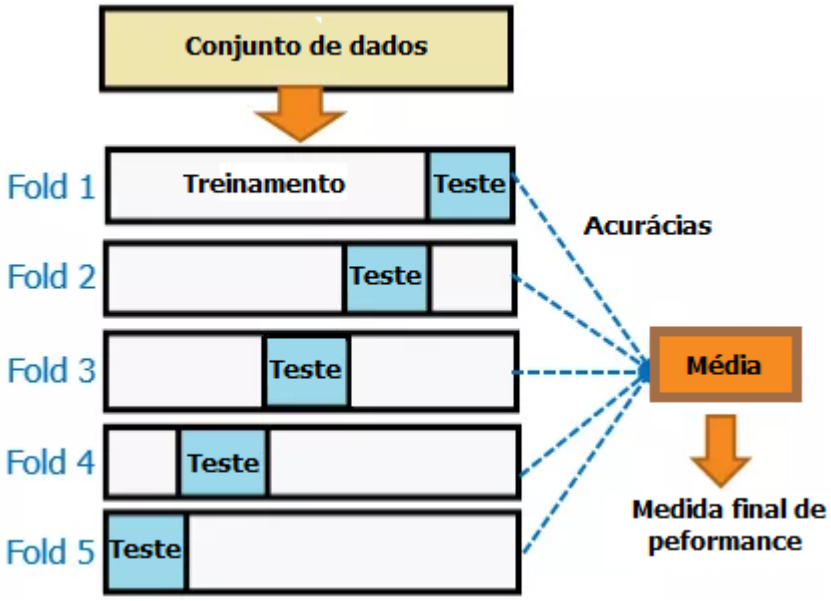


Figura 10: Algoritmo de validação cruzada *k-fold* para k = 5. Adaptada de (Srinidhi, 2018)

2.6.4.Matriz de confusão

No campo da aprendizagem de máquinas, uma matriz de confusão é uma tabela que é frequentemente usada para descrever o desempenho de um modelo de classificação sobre um conjunto de dados de teste para os quais os valores verdadeiros são conhecidos. Também conhecida como matriz de erros, ela permite a fácil identificação da confusão entre classes, por exemplo, quando os exemplos de uma classe são rotulados erroneamente como outra pelo classificador. A maioria das medidas de desempenho é computada a partir de matrizes de confusão (TING, 2011).

Uma matriz de confusão é um resumo dos resultados de previsão sobre um problema de classificação. O número de previsões corretas e incorretas é resumido com valores de contagem e dividido por cada classe. Esta é a chave para a matriz de confusão. Ela nos dá uma visão não apenas dos erros que estão sendo cometidos por um classificador, mas, mais importante, dos tipos de erros que estão sendo cometidos. Uma matriz de confusão é exemplificada na Figura 11.

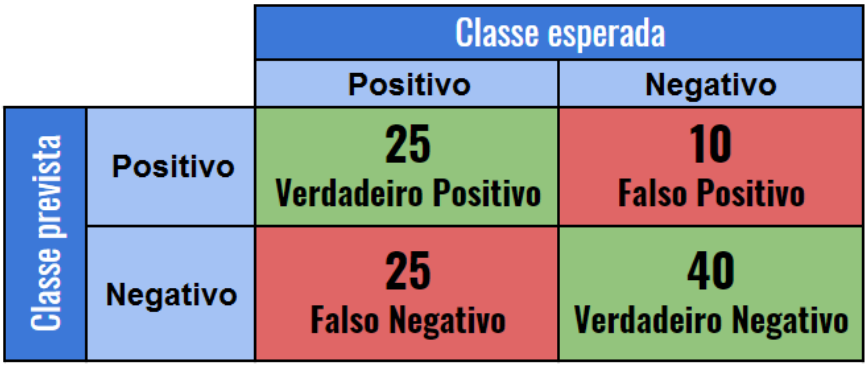


Figura 11: Matriz de confusão com resultados fictícios.

Como pode ser visto na Figura 11, a relação entre a classe prevista e a classe esperada de um exemplo define se ele foi classificado como verdadeiro positivo (ou "VP"), falso positivo (ou "FP"), verdadeiro negativo (ou "VN") ou falso negativo (ou "FN"). A partir desses resultados, na Tabela 4 são definidas algumas métricas para o desempenho dos classificadores (DANGETI, 2017) (TING, 2011).

Tabela 4: Métricas para o desempenho dos classificadores. As variáveis são o número de casos verdadeiros positivos (ou "VP"), falsos positivos (ou "FP"), verdadeiros negativos (ou "VN") ou falsos negativos (ou "FN").

Métrica	Fórmula	Significado
Acurácia	$\frac{VP + VN}{VP + VN + FP + FN}$	A proporção de exemplos preditos corretamente sobre o total de exemplos. Mostra o desempenho geral do modelo de classificação.
Sensibilidade (s)	$\frac{VP}{VP + FN}$	A capacidade do classificador em predizer uma condição positiva para casos que realmente a têm. Pode ser calculada para cada classe do problema.
Precisão (p)	$\frac{VP}{VP + FP}$	Razão entre as predições positivas corretas e o número total de predições positivas. Pode ser calculada para cada classe do problema.
Coeficiente F1	$2 \cdot \frac{s \cdot p}{s + p}$	Métrica de desempenho com alta confiabilidade que considera tanto a precisão como a sensibilidade. Pode ser calculada para cada classe do problema.

2.6.5. Máquina de Vetores de Suporte

Como exemplo de algoritmo de aprendizado de máquinas, a Máquina de Vetores de Suporte (do inglês, "*Support Vector Machine*", ou "SVM") é um algoritmo originalmente aplicado para problemas de classificação binária e tem se elevado como um dos classificadores mais populares em vários campos.

A base deste método é a definição de um hiperplano no espaço de descrição dos exemplos a serem classificados de forma separar linearmente as diferentes classes do problema. Se os dados do problema não forem linearmente separáveis na dimensão original do espaço de descrição, deve ser feita a seleção de uma função, chamada núcleo (do inglês, "*kernel function*"). Essa função mapeia o espaço de descrição em um espaço de dimensão mais elevada, na qual os dados sejam, efetivamente, linearmente separáveis por um hiperplano (LORENA, 2003). Portanto, num espaço dimensional superior, uma classificação linear se torna possível, como pode ser visto na Figura 12 (WANG e GUO, 2014).

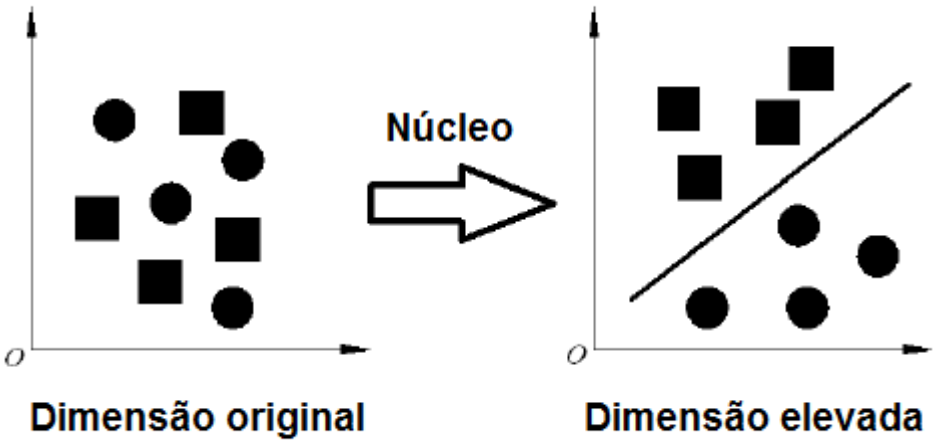


Figura 12: Ilustração do processo de mapeamento do espaço de descrição realizado por classificadores SVM. Adaptado de (WANG e CHO, 2014).

Então, um hiperplano é parametrizado e movido entre as classes até deixar a margem máxima possível, como pode ser visto na Figura 13 (WANG e GUO, 2014). Assim, o princípio deste algoritmo é determinar um hiperplano ótimo através de um conjunto de amostras de treinamento, tornando-o um classificador linear binário não-probabilístico (LORENA, 2003).

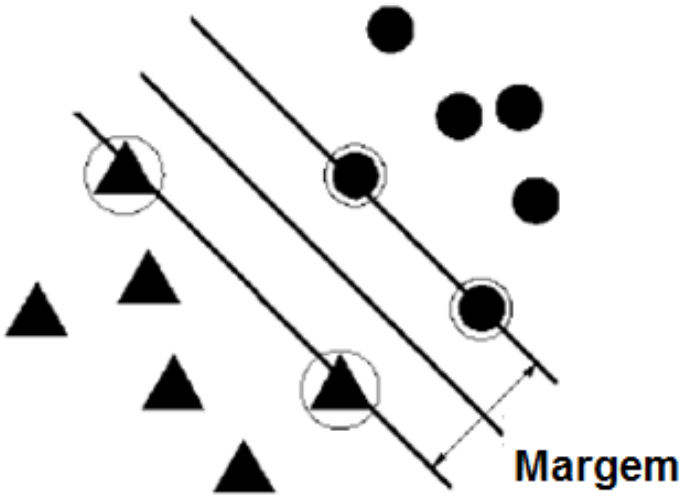


Figura 13: Ilustração da parametrização de um hiperplano ótimo com margem máxima por um algoritmo SVM. Adaptado de (WANG e CHO, 2014).

Dessa forma, fundamentado na teoria de aprendizagem estatística, o SVM foi desenvolvido com o objetivo de ser uma ferramenta capaz de reconhecer padrões na distribuição de dados, tendo como principal característica robustez diante de um conjunto de informações de grandes dimensões, além de ótima capacidade de generalização. A teoria de aprendizagem estatística (TAE), base do SVM, visa estabelecer condições matemáticas que permitam a escolha de um classificador f' com bom desempenho para conjuntos de treinamento e teste. Ou seja, busca-se uma função f' capaz de classificar os dados de treinamento da forma mais correta possível, sem dar atenção exacerbada a qualquer ponto individual do mesmo. Por exemplo, seja f um classificador para o espaço de classes e, F um conjunto de todos classificadores que um determinado algoritmo pode gerar. Este algoritmo, durante o aprendizado, utiliza um conjunto de treinamento S, composto por n pares de dados (xi,yi), em que yi representa a classe do padrão xi para gerar um classificador particular $f' \in F$ (LORENA, 2003).

Além disso, o SVM é reconhecido como um classificador robusto que maximiza a precisão preditiva, minimizando o sobreajuste (CHO e LEE, 2013).

Entre os possíveis núcleos a serem aplicados no processo de classificação por algoritmos SVM, a Tabela 5 expõe algumas das funções mais comuns e suas principais características para um vetor de atributos x e um vetor de parâmetros z (AMID e AGHDAM e AMINDAVAR, 2012). É importante ressaltar que os núcleos estão ordenados por nível de complexidade, a qual é um fator fundamental a ser levado em conta no processo de escolha do núcleo adequado para minimizar erros de sobreajuste ou sub-ajuste.

Tabela 5: Funções núcleo mais comuns (AMID e AGHDAM e AMINDAVAR, 2012).

Núcleo	Equação	Descrição
Linear	$\kappa_l(x,z) = x \cdot z$	Usado para um espaço de descrição linearmente separável em sua dimensão original.
Quadrático	$\kappa_q(x,z) = x \cdot z(1 + x \cdot z)$	Uma função simples e eficiente utilizada para muitas aplicações.
Polinomial	$\kappa_q(x,z) = (\langle x,z\rangle + R)^d$	R e d são parâmetros do núcleo. É intuitivo esperar que o peso relativo do polinômio de ordem mais alta aumente diminuindo R.
Gaussiano	$\kappa_q(x,z) = e^{(-\frac{\|x-z\|^2}{2\sigma^2})}$	O parâmetro σ controla a flexibilidade do núcleo. Portanto, um pequeno σ aumenta o risco de sobreajuste enquanto um grande σ transforma o núcleo em uma função constante que não é utilizável para treinar qualquer classificador relevante.

Quando mais de duas classes são necessárias, é preciso aplicar um método multiclasse, já que o algoritmo básico do SVM é binário, ou seja, só aceita duas classes possíveis na saída. Alguns dos principais métodos multiclasse são conhecidos como: Um contra todos (OAR), Um contra um (OAO), máquina de vetores de suporte multiclasse direta (DMSVM) e método de árvore binária de máquina de vetores de suporte multiclasse (WANG e CHO, 2014).

2.7. Preparação dos dados

Uma etapa fundamental para a modelagem de algoritmos de classificação por aprendizado de máquina é a preparação dos dados a serem utilizados como entrada para os algoritmos de treinamento. Essa etapa também é conhecida como pré-processamento.

Essa etapa é indispensável, pois ela permite, primeiramente, compreender o comportamento dos dados a serem analisados, um conhecimento fundamental para, em seguida, definir o algoritmo de treinamento adequado. Além disso, ela também tem o papel de gerar transformações nos dados que aumentem a eficiência dos algoritmos de classificação.

Nas próximas seções serão apresentados três estágios importantes para a preparação de dados: o balanceamento de classes, fundamental para ajustar a distribuição de classes no conjunto de dados; as análises estatísticas, realizadas com o objetivo de explorar as propriedades estatísticas dos exemplos e serem classificados e seus atributos; e a seleção de atributos, realizada com o objetivo de determinar qual é o conjunto mínimo de atributos dentre os disponíveis para descrever os exemplos de forma satisfatória.

2.7.1. Balanceamento de classes

Muitos aspectos podem influenciar negativamente o desempenho de um modelo de classificação criado por um sistema de aprendizado supervisionado. Um desses aspectos é chamado de desbalanceamento de classes, definido como um conjunto de dados no qual as classes não estão igualmente representadas, isto é, quando as classes tem diferentes números de exemplos. Comumente, pequenas variações da distribuição de exemplos por classe não geram efeitos consideráveis no desempenho dos classificadores, porém, quando essa diferença é grande, os sistemas de aprendizado tendem a valorizar classes com maior número de exemplos (majoritárias) em detrimento às classes com menor número de exemplos (minoritárias).

Esse problema é recorrente no aprendizado de máquina, pois em muitos problemas reais há naturalmente uma grande desproporção no número de casos pertencentes a cada uma das classes. Por exemplo, na detecção de doenças raras em conjuntos de dados médicos, na detecção de palavras específicas em conjuntos de classificação de texto, entre outros. Além disso, em geral, os exemplos da classe

minoritária são aqueles em que se tem o maior interesse de obter classificações corretas (SANTOS, 2016).

Muitos métodos podem ser aplicados a conjuntos de dados para solucionar o problema de classes desbalanceadas, chamados métodos de balanceamento de classes. Esses podem ser basicamente divididos em duas classes de métodos, ilustrados na Figura 14.

• Métodos de sub-amostragem (do inglês, "*under-sampling*") que tem como objetivo balancear o conjunto de dados pela eliminação de exemplos da classe majoritária;

• Métodos de sobre-amostragem (do inglês, "*over-sampling*") que criam cópias dos exemplos da classe minoritária com o objetivo de obter uma distribuição balanceada com a classe majoritária.

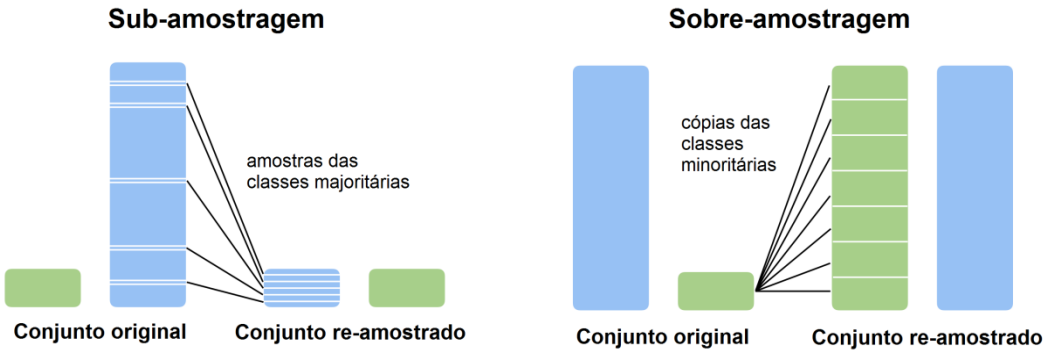


Figura 14: Ilustração dos métodos de balanceamento de classes.

Ambas as classes de métodos se baseiam no mesmo princípio: incorporar um viés no conjunto de dados para selecionar mais exemplos de certas classes do que de outras como forma de compensar o desbalanceamento presente nos dados. Por essa razão, ambos têm problemas correlacionados. A sub-amostragem pode eliminar exemplos potencialmente úteis nos dados e a sobre-amostragem pode aumentar a probabilidade de ocorrer um sobreajuste, uma vez que a maioria dos métodos de sobre-amostragem fazem cópias exatas dos exemplos da classe minoritária (SANTOS, 2016).

2.7.2. Análise estatística dos dados

Com objetivo de explorar mais profundamente o comportamento estatístico dos dados analisados e seus atributos, diversos tipos de análise de inferência estatística podem ser realizados (BOSLAUGH, 2012). Entre esses, três classes importantes de

testes serão apresentadas: os testes de normalidade, os testes de correlação de atributos e testes de hipótese sobre a média dos dados.

- Testes de normalidade:

Dentro do campo da inferência estatística, existe uma classe de testes de hipótese utilizados para deduzir se a distribuição estatística de certa amostra de dados pode ser aproximada por uma distribuição normal. Eles são importantes para determinar se, uma vez que apresentem normalidade, os dados admitem o uso de testes paramétricos para o prosseguimento das análises estatísticas exploratórias (BOSLAUGH, 2012). Na Tabela 6, são apresentados três testes de normalidade convencionais implementados na biblioteca do software comercial MATLAB (MATHWORKS, 2019).

Tabela 6: Testes de normalidade implementados no software MATLAB.

Nome	Descrição
Teste de Lilliefors (TLF)	Testa a hipótese nula de que os dados analisados advêm de uma população com distribuição normal contra a alternativa de não advir de tal distribuição, sem especificar sua esperança nem variância. O resultado do teste é 1 se o teste rejeitar a hipótese nula, e 0 caso contrário. O teste de Lilliefors é adequado quando os parâmetros da distribuição nula são desconhecidos e devem ser estimados (MATHWORKS, 2019).
Teste de Anderson-Darling (TAD)	Com mesmas hipóteses que o teste de Lillifors, o teste Anderson-Darling é igualmente usado para testar se uma amostra de dados vem de uma distribuição normal. A estatística do teste pertence à família das estatísticas da função de distribuição empírica quadrática (MATHWORKS, 2019).
Teste de Jarque-Bera (TJB)	Igualmente análogo às hipóteses dos dois testes anteriores, o teste de Jarque-Bera também não requer a especificação dos parâmetros da distribuição nula. O teste é projetado especificamente para alternativas no sistema de distribuições de Pearson. Para amostras grandes, a estatística do teste tem uma distribuição qui-quadrada com dois graus de liberdade (MATHWORKS, 2019).

- Testes de correlação:

Além dos testes de normalidade, uma segunda classe de testes de hipótese comuns na preparação de dados são os testes de correlação de atributos. Esses testes buscam determinar se existe alguma associação entre os atributos do conjunto de exemplos estudados. Essa determinação é fundamental para compreender a relação entre os atributos disponíveis para expressar os exemplos, e se algum atributo, por ser altamente correlacionado com os outros, apenas carrega informação redundante sem relevância estatística. De fato, a eliminação de atributos redundantes é um dos objetivos principais do processo de preparação de dados, e será mais profundamente explorada na próxima seção dessa revisão dada sua relevância. (BOSLAUGH, 2012). Na Tabela 7, são apresentados três testes de correlação convencionais implementados na biblioteca do software comercial MATLAB (MATHWORKS, 2019).

Tabela 7: Testes de correlação implementados no software MATLAB.

Nome	Descrição
Teste linear de Pearson (TLP)	O teste de correlação de Pearson é o teste paramétrico de correlação de atributos mais comum. Ele determina um coeficiente entre -1 e 1 que determina a relação linear entre dois atributos de um conjunto de exemplos. Se os atributos testados apresentam uma forte relação linear direta o coeficiente se aproxima de 1, se apresentam uma forte relação linear reversa o coeficiente se aproxima de -1, e se os atributos não apresentam relação linear, o coeficiente se aproxima de 0.
Teste dos postos de Spearman (TPS)	A correlação de Spearman é um teste não-paramétrico (não requer normalidade) que mede como um atributo aumenta enquanto outro atributo também aumenta, sem exigir que o aumento seja representado por uma relação linear (MATHWORKS, 2019).
Teste dos postos de Kendall (TPK)	Similar ao teste de Spearman, o teste de Kendall é um teste não-paramétrico igualmente baseado apenas na relação ordinal entre os atributos (MATHWORKS, 2019)..

- Testes de sobreposição (ou de hipótese da média):

Além dos testes de normalidade e de correlação, outra análise importante para compreender o comportamento dos dados e seus atributos são testes de hipótese para determinar se os exemplos rotulados com classes diferentes pertencem ou não a uma mesma população estatística. Esse resultado é importante para compreender se os atributos utilizados para expressar os exemplos são capazes de separar efetivamente os dados nas classes estudadas. Caso os testes falhem, conclui-se que, apesar de dois conjuntos de exemplos serem previamente rotulados com classes diferentes, o atributo testado é estatisticamente incapaz de diferenciar os dois conjuntos (BOSLAUGH, 2012). Na Tabela 8, são apresentados dois testes de hipótese convencionais implementados na biblioteca do software comercial MATLAB (MATHWORKS, 2019).

Tabela 8: Testes de hipótese da média implementados no software MATLAB.

Nome	Descrição
Teste dos postos de Wilcoxon (TPW)	Testa a hipótese nula que, para um determinado atributo, os exemplos de um par de classes diferentes provêm de uma mesma distribuição contínua de medianas iguais, contra a hipótese que eles não provêm. O teste assume que os dados das duas amostras são independentes.
Teste de Kolmogorov-Smirnov de duas amostras (TKS)	Testa a hipótese nula que, para um determinado atributo, os exemplos de um par de classes diferentes provêm de uma mesma distribuição contínua, contra a hipótese que eles não provêm. O teste se baseia na avaliação da diferença entre as funções de distribuição cumulativas de cada classe.

2.7.3. Seleção de atributos

O desafio fundamental de quase todas as tarefas de modelagem de dados é identificar e caracterizar as relações entre um ou mais atributos dos dados e suas

classes. Na maioria dos conjuntos de dados, apenas um subconjunto de atributos disponíveis são características relevantes, ou seja, realmente informativas na determinação do valor das classes. Os restantes são chamados de atributos irrelevantes, ou não informativos, os quais são raramente distinguíveis a priori nos problemas do mundo real (URBANOWICZ ET AL., 2017).

Um atributo é chamado irrelevante quando existe um subconjunto do universo de atributos disponíveis, do qual ele não faz parte, que descreve de forma completa e consistente as classes a serem aprendidas. Uma descrição das classes é dita completa se é capaz de classificar todos os exemplos, e consistente se classifica corretamente os exemplos, como é ilustrado na Figura 15. Atributos irrelevantes também podem ser chamados de redundantes se existe alguma combinação linear de outros atributos que tenha uma alta correlação com ele. Ou seja, em outras palavras, se toda a informação útil nesse atributo já é fornecida por outros.

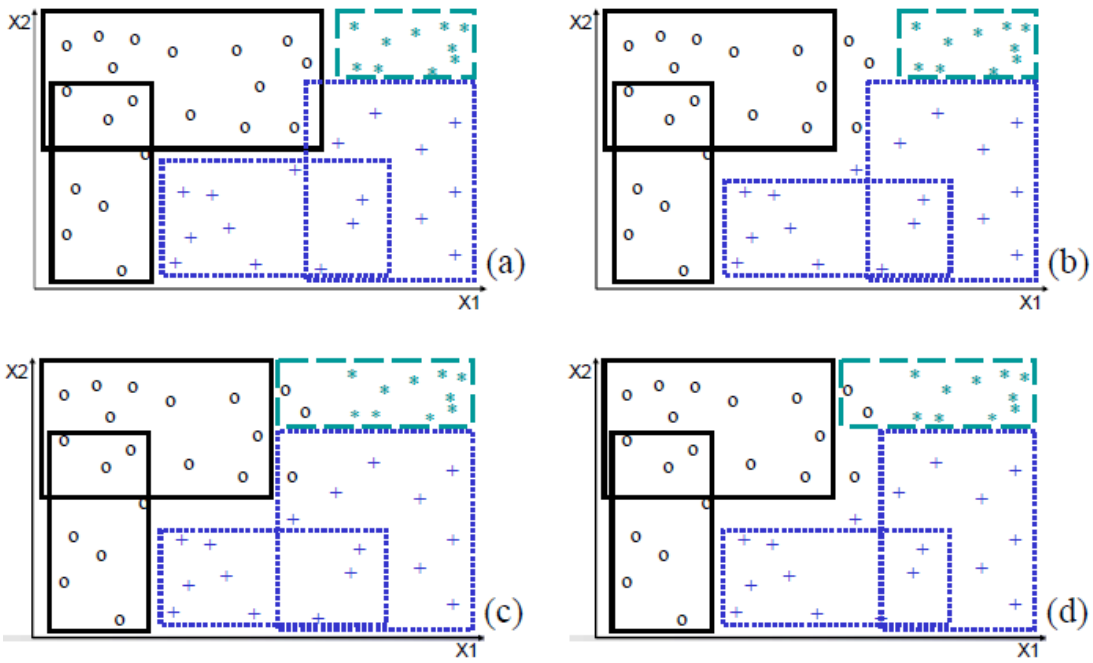


Figura 15: Relação entre completude e consistência da classificação de um espaço de descrição bidimensional com atributos X1 e X2. (a) Descrição completa e consistente; (b) Incompleta e consistente; (c) Completa e inconsistente; (d) Incompleta e inconsistente. Adaptada de (REZENDE, 2003).

A presença de atributos irrelevantes na descrição dos exemplos traz uma série de prejuízos para a modelagem de algoritmos classificadores. O primeiro é o fato da sua presença aumentar a dimensionalidade do problema, a qual aumenta a complexidade do modelo de classificação, aumenta o tempo de processamento do método de

treinamento; e torna o espaço de descrição mais esparso, aumentando a probabilidade de sobreajuste (URBANOWICZ ET AL., 2017) (SPRUYT, 2014).

Por essa razão, métodos de seleção de atributos podem ser definidos genericamente como o processo de identificação de atributos relevantes e descartando os irrelevantes. Esses métodos podem ter três abordagens diferentes (URBANOWICZ ET AL., 2017):

- Filtros (do inglês, "*filters*"): Métodos que definem uma ou mais métricas calculadas a partir dos atributos gerais dos exemplos de treinamento para determinar os atributos relevantes antes do processo de treinamento. Os filtros são geralmente muito mais rápidos computacionalmente e funcionam independentemente do algoritmo de treinamento utilizado. Os filtros podem ser classificados pelas métricas que empregam, por exemplo, distância, dependência, consistência, similaridade, estatísticas medidas, entre outras.

- *Wrappers:* Métodos que realizam o treinamento e a validação de um determinado classificador para subconjuntos do universo de atributos disponíveis até determinar o subconjunto de melhor desempenho. O método pode testar todos os subconjuntos de atributos possíveis ou uma parcela deles, geralmente utilizando algoritmos de busca auxiliares. Por essa razão, os *wrappers* são métodos muito eficientes, porém muito intensivos computacionalmente, fato que, em geral, limita seu uso apenas para modelos de classificação de baixa complexidade.

- Métodos agregados (do inglês, "*embedded methods*"): Métodos que incorporam o processo de seleção ao processo de execução dos algoritmos de classificação. Esses métodos tendem a ser computacionalmente mais eficientes que os *wrappers*, porém também são dependentes do classificador utilizado, o que os diferenciam dos filtros.

Uma família de algoritmos de seleção do tipo filtro muito comum é chamada de *Relief*. A ideia central dos algoritmos da família *Relief* é estimar a qualidade dos atributos com base no quão bem o atributo pode distinguir entre exemplos que estão próximos uns dos outros no espaço de descrição. Os algoritmos mais comuns dessa família são (URBANOWICZ ET AL., 2017):

(1) *Relief* básico: restrito apenas para problemas de classificação binária;

(2) *ReliefF*: extensão do *Relief* básico para problemas com mais de duas classes;

(3) *RReliefF*: *Relief* adaptado para problemas de regressão.

Uma grande vantagem desses algoritmos é que eles não realizam a seleção através de combinações de atributos, mas sim utilizam o conceito de vizinhos mais próximos para obter estatísticas de atributos. Essas estatísticas indiretamente consideram interações entre os atributos analisados, o que é uma vantagem usualmente associada aos *wrappers* e métodos agregados (URBANOWICZ ET AL., 2017).

Na prática, os algoritmos *Relief* calculam um coeficiente de relevância (ou peso) para cada atributo analisado entre -1 e 1 (onde 1 é a maior relevância), avaliando assim os atributos em ordem crescente de relevância. Dessa forma, é possível determinar quais atributos apresentem uma relevância baixa, podendo ser excluídos do modelo de classificação (URBANOWICZ ET AL., 2017).

3. METODOLOGIA

3.1. Materiais analisados

Os dados de treinamento usados como base para a classificação foram obtidos pela inspeção de segmentos de tubos de aço HP-Nb de 100mm (Figura 16) retirados de fornos de reforma a vapor industrial com duas composições químicas diferentes, que chamaremos de A e B, conforme a Tabela 9.

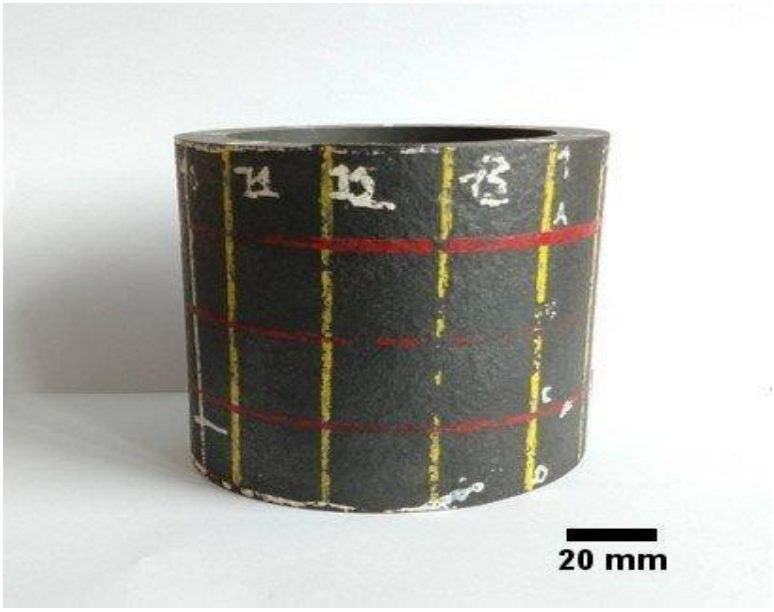


Figura 16: Fotografia de uma das amostras.

Tabela 9: Composições químicas dos grupos de amostras analisadas (% em peso).

Grupo de amostras	Ni	Cr	C	Nb	Ti	Si	Mn	P	Zr	S	Mo	Co	W
A	35,0	25,5	0,54	1,13	0,083	1,60	1,30	0,020	0,001	0,006	<0,01	<0,02	0,039
B	34,2	25,8	0,46	1,18	<0,01	1,94	1,24	0,023	<0,001	0,011	0,034	0,10	0,057

Cada segmento foi retirado de uma determinada região das colunas onde, de acordo com as condições de temperatura, pressão e tensão às quais a região estava

exposta, se esperava ter um mesmo estado de evolução microestrutural em todo o segmento. Isso é necessário para garantir que todos os dados captados de cada segmento sejam garantidamente associados a uma determinada classe. A validação do estado integral esperado para cada segmento foi feita através de ensaios de metalografia realizados em parceria com o Laboratório de Propriedades Mecânicas (PROPMEC/UFRJ). Portanto, buscou-se selecionar uma amostra para cada estado, os quais são seis, para cada uma das duas composições, as quais são duas, totalizando doze amostras necessárias. Todavia, para o grupo B não foi possível extrair amostras para o estado VI, totalizando então onze amostras, como mostrado na Tabela 10.

Tabela 10: Amostras selecionadas para cada grupo de amostras.

Grupos de amostras	Amostras por estado de evolução microestrutural
A	I, II, III, IV, V, VI
B	I, II, III, IV, V

3.2. Coleta de dados

Na coleta de dados foi utilizado sistema de correntes parasitas desenvolvido no Laboratório de Ensaios Não-Destrutivos, Corrosão e Soldagem (LNDC). O sistema é constituído por dois sensores: um sensor de correntes parasitas convencional (sensor 1) e um sensor com magnetização externa (sensor 2), além de eletrônica de excitação e obtenção de sinal desenvolvida pelo próprio laboratório, constituindo um módulo denominado Sonda Híbrida. A sonda foi projetada por (ARENAS, 2017), levando em consideração a geometria dos tubos garantindo assim o acoplamento entre eles. A sonda é movimentado por um braço robótico automático, tecnologia que permite um posicionamento superficial eficiente ao longo do comprimento da amostra, além de possibilitar um movimento de velocidade constante sobre a mesma.

Além das inspeções das onze amostras que pertencem cada uma a um estado de evolução microestrutural diferente, foi captado também o sinal correspondente a passagem da sonda híbrida ao ar. Os dados associados a esse sinal são importantes pois podem ser utilizados para indicar o estado no qual a sonda tenha se desprendido do material analisado durante uma inspeção em campo, o que é prejudicial para o funcionamento de um sistema automático. Por essa razão, esses dados serão incorporados ao modelo de classificação proposto nesse trabalho como uma classe

adicional, chamada de classe 0, ou AR. Na Tabela 11 estão, portanto, enumeradas as classes do modelo de classificação proposto.

Tabela 11: Classes do modelo de classificação proposto.

Classes do modelo de classificação	Significado físico
0/AR	Sinal da sonda ao ar
1	Sinal do estado I
2	Sinal do estado II
3	Sinal do estado III
4	Sinal do estado IV
5	Sinal do estado V
6	Sinal do estado VI

Por fim, tem-se o computador que carrega todo o sistema de armazenamento e tratamento dos dados via o software MATLAB®. Após serem carregados, os sinais adquiridos passam por um algoritmo que realiza o cálculo e extração de três variáveis que servirão de atributos para o treinamento do modelo de classificação: a amplitude, a fase e o offset. Essas três variáveis são calculadas para cada um dos dois sensores que constituem a Sonda Híbrida, resultando num total de seis atributos, como mostra a Tabela 12:

Tabela 12: Atributos dos exemplos do modelo de classificação proposto.

Atributos	A(EC)	F(EC)	O(EC)	A(SF)	F(SF)	O(SF)
Significado físico	Amplitude sensor 1	Fase sensor 1	Offset sensor 1	Amplitude sensor 2	Fase sensor 2	Offset sensor 2

3.3. Análise de compatibilidade de classes

Antes de prosseguir para o pré-processamento dos dados, foi realizada uma análise para determinar se as distribuições estatísticas dos dados associados aos mesmos estado de evolução microestrutural correspondentes às amostras das duas composições químicas distintas são estatisticamente compatíveis entre si. Em outras

palavras, determinar se os dados de um mesmo estado de evolução das duas composições diferentes pertencem à mesma população estatística.

Para isso, foi realizada uma análise de sobreposição utilizando os testes TPW e TKS, com nível de significância de 5%, que busca determinar se os pontos correspondentes aos mesmos estados de evolução microestrutural das amostras das duas diferentes composições provêm de distribuições estatísticas de medianas iguais. Para isso, os testes foram aplicados às distribuições referentes a cada um dos atributos, isto é, para os dados obtidos de cada variável experimental da sonda híbrida.

3.4. Pré-processamento

No fim da etapa anterior obtivemos um conjunto de exemplos constituídos de vetores de seis atributos, cada exemplo representando um ponto escaneado pela Sonda Híbrida durante a inspeção das amostras. Esse conjunto é dividido em subconjuntos de seis classes para o grupo de amostras A, cinco classes para o grupo de amostras B, e um subconjunto de exemplos da classe 0. O número de exemplos captados nas inpeções de cada subconjunto é mostrado na Tabela 13.

Tabela 13. Número de pontos de cada classe para cada fonte dos dados.

Classes	Pontos do grupo de amostras A	Pontos do grupo de amostras B	Pontos da sonda ao ar
0	0	0	1044
1	1547	6160	0
2	2145	3728	0
3	1625	6432	0
4	2132	2976	0
5	2431	4544	0
6	2135	0	0

Como pode ser visto na Tabela 11, devido a variações experimentais no processo de inspeção, o número de exemplos de cada subconjunto é muito discrepante, o que prejudica o processo de treinamento. Para garantir o balanceamento das classes de cada conjunto de amostras, um método de balanceamento por sub-amostragem foi aplicado. Nele, selecionou-se aleatoriamente 1000 (mil) exemplos de cada subconjunto utilizando uma rotina no software MATLAB®. Com os exemplos selecionados, foram compostos três novos conjuntos de dados: o conjunto A_P, composto pelos exemplos

amostrados provenientes dos dados captados do grupo de amostras A mais os pontos da sonda ao ar, e o conjunto B_P, composto pelos exemplos amostrados provenientes dos dados captados do grupo de amostras B mais os pontos da sonda ao ar.

Após o balanceamento, o conjunto total de exemplos (chamado AB_P), foi submetido às análises estatísticas de aprimoramento do modelo, como mostrado na revisão teórica. Para todas os testes estatísticos, foi considerado o valor de $\alpha = 0.05$ para o nível de significância. As análises foram compostas por:

a. Análise de normalidade: Constituída pelos testes TLF, TAD e TJB. Para os pontos de cada classe do conjunto, um teste de normalidade foi realizado para a distribuição de cada um de seus atributos. Foi utilizado como critério de normalidade que ao menos um dos testes acima acusasse normalidade.

b. Análise de correlação: Constituída pelos testes TLP, TPS e TPK. Os dados de cada par de atributos do conjunto tiveram sua correlação medida por um coeficiente de correlação. Foi utilizado como critério de correlação que ao menos um teste acima acusasse um coeficiente de correlação de módulo acima de 0,7.

c. Análise de sobreposição: Constituída pelos testes TPW e TKS. Para cada atributo do problema, foram selecionados pares de classes do conjunto, testando a hipótese de que suas medianas sejam iguais, isto é, de que elas estariam sobrepostas. Nesse caso, considera-se que o atributo é incapaz de diferenciar o par de classes testadas. Foi utilizado como critério de sobreposição que ao menos um dos testes acusasse a sobreposição do par de classes.

Ao fim das análises estatísticas, o conjunto AB_P passará por uma análise de seleção de atributos, utilizando o método RELIEF-F, implementado no software MATLAB®. Esse método gerará como resultado um ordenamento de relevância entre os atributos dos exemplos, fornecendo para cada atributo um coeficiente de relevância. Com esse resultado em mãos, foram selecionados os atributos considerados irrelevantes, e, ao final do estágio de treinamento, a influência da exclusão deles nas acurácias dos modelos será testada.

Chegando ao fim da etapa de pré-processamento, foram obtidas versões processadas dos conjuntos de dados iniciais: A_P, B_P, e um conjunto total processado,

composto pela união desses dois conjuntos, AB_P. Esses três conjuntos serão aqueles efetivamente utilizados para o processo de treinamento do modelo de classificação.

3.5. Visualização dos dados

Antes de dar início ao processo de treinamento, foi realizada uma análise visual da distribuição dos dados utilizando a metodologia de gráfico de coordenadas paralelas. Essa metodologia é importante para disponibilizar uma forma de visualizar os dados multidimensionais de forma prática e compreensível durante o processo de modelagem, sem depender apenas de indicadores numéricos. Um exemplo desse tipo de gráfico pode ser observado na Figura 17.

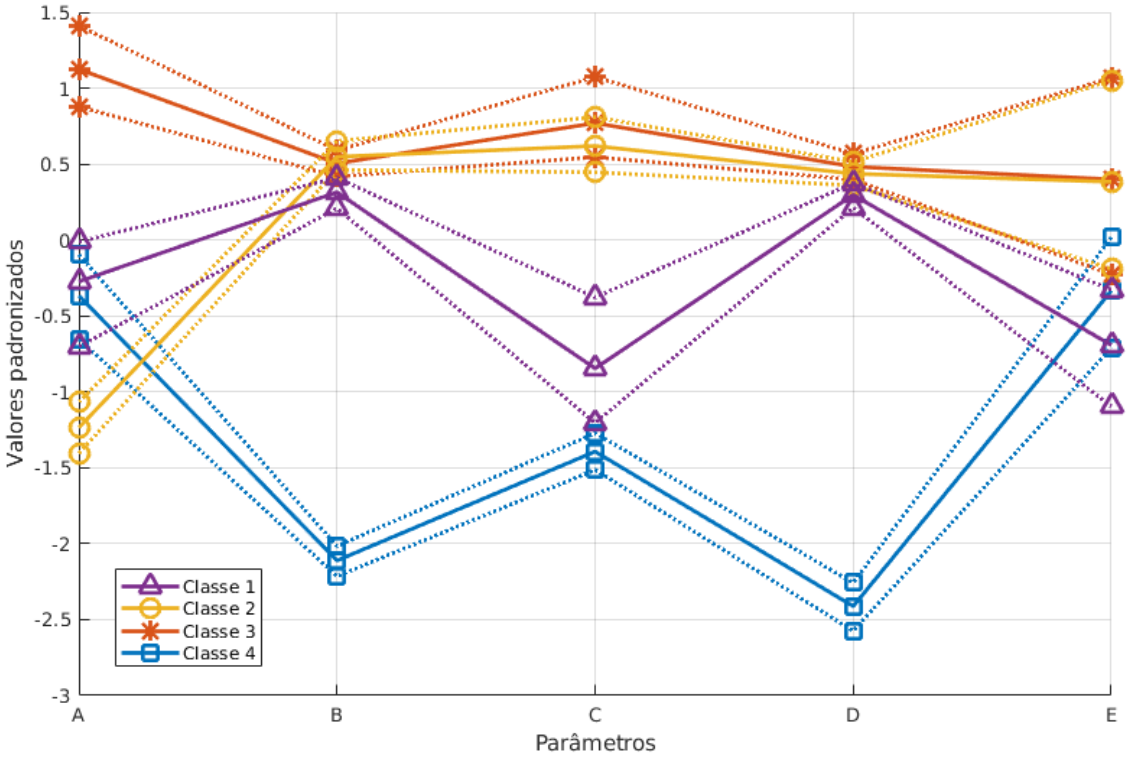


Figura 17: Exemplo de gráfico de coordenadas paralelas.

Gráficos de coordenadas paralelas são ferramentas muito úteis para a visualização de dados em dimensão muito elevada (como no caso atual, em seis dimensões) em apenas duas dimensões, sem gerar uma perda de informação e sem depender de métodos matemáticos específicos como a análise de componentes principais, por exemplo.

Esses gráficos são construídos representando cada exemplo como uma linha que une os valores de cada atributo em seu respectivo eixo. Em conjuntos com número

muito elevado de exemplos, como o deste trabalho, utiliza-se a estratégia de dividir os exemplos em suas classes, representando apenas as linhas correspondentes às médias de cada classe (em linha cheia) e, quando possível, os quartis de 10% (em linha tracejada), como pode ser visto na Figura 17. No trabalho atual, entretanto, as linhas dos quartis foram suprimidas para facilitar a visualização. Além disso, os valores de cada eixo são usualmente normalizados ou padronizados para garantir uma igualdade nas escalas de cada um.

Dos dados processados, três gráficos foram analisados: a) os dados do conjunto A_P; b) os dados do conjunto B_P; e; c) os dados do conjunto AB_P.

3.6. Treinamento e validação do modelo de classificação

Após a análise visual, os dados estão prontos para o processo de treinamento utilizando o modelo de SVM quadrático. A seleção desse modelo, assim como o refino de seus parâmetros foi realizado em trabalhos anteriores (MARQUES, 2019) (DUTRA, 2019) utilizando o mesmo conjunto de dados.

Os dados serão utilizados para treinar três modelos, cujos desempenhos de classificação serão comparados. Os primeiros dois modelos levam em consideração apenas os dados de cada uma das duas composições químicas mais o sinal da sonda ao ar (modelos MA e MB), e o terceiro modelo leva em consideração todos os dados (MAB).

Após o treinamento, será realizada a validação do desempenho de cada modelo na classificação do seus próprios conjuntos de treinamento. Essa validação será avaliada a partir de suas matrizes de confusão, seus valores de acurácia e seus coeficientes F1 calculados para cada classe. O objetivo dessa validação será mostrar se o modelo MAB terá uma acurácia satisfatória e um baixo valor de erro cometido na classificação de cada estado de evolução microstrutural em comparação aos modelos MA e MB.

Depois disso, será realizada uma etapa de teste, na qual será calculada a acurácia de cada modelo na classificação de todos os dados, isto é, do conjunto AB_P. Ou seja, cada modelo será usado para prever a classificação de todos os dados vindos das duas composições distintas simultaneamente. Esse teste terá o objetivo de determinar se o modelo MAB terá um desempenho de classificação mais elevado que os modelos MA e MB. Em caso positivo, será possível concluir que a metodologia proposta foi capaz de gerar um modelo geral para as duas composições que se comporta melhor que os modelos individuais no momento de classificar dados com composição variável.

A Tabela 14 mostra um resumo dos três modelos e os conjuntos utilizados durante seu treinamento, validação e teste.

Tabela 14: Resumo dos três modelos.

Modelo	Algoritmo de treinamento	Conjunto de treinamento	Validação	Conjunto de teste
MA	SVM com núcleo quadrático	A_P	Validação cruzada com k = 10	AB_P
MB		B_P		
MAB		AB_P		

3.7. Avaliação da seleção de atributos

Por fim, os modelos MA, MB e MAB serão treinados novamente, porém removendo os atributos indicados como irrelevantes pela análise estatística realizada anteriormente. Após cada treinamento, será calculada a acurácia de cada novo modelo, que serão posteriormente comparados. Serão realizados três novos treinamentos:

1. Sem o atributo O(SF);
2. Sem o atributo A(SF);
3. Sem os atributos O(SF) e A(SF);

Se for determinado que a retirada de algum atributo teve pouco impacto na acurácia de algum dos modelos, será proposta a exclusão desse atributo do conjunto ideal de atributos para os modelos.

4. RESULTADOS E DISCUSSÃO

4.1. Análise de compatibilidade de classes

Os resultados da análise de compatibilidade de classes mostraram que apenas o estado III para os dados referentes ao atributo O(SF) apresentou compatibilidade estatística entre as duas composições. Isso significa que, de fato, os testes aplicados indicam que os dados referentes às duas diferentes composições provém de populações diferentes.

Esse resultado é importante para mostrar que a alta variabilidade entre os dados provenientes de cada uma das duas composições de fato torna sua classificação simultânea em um modelo único um problema não trivial que requer uma modelagem computacional e estatística adequada.

4.2. Pré-processamento

4.2.1. Análise de normalidade

Os resultados da análise de normalidade podem ser observados na Figura 18. Observa-se que apenas 11 das 46 distribuições estatísticas de cada classe para os dados de cada atributo foram considerados provindos de uma população normal. Isso é um forte indício de não-normalidade no problema, justificando, portanto a escolha de testes não paramétricos nas seguintes etapas desse estudo estatístico.

4.2.2. Análise de correlação

Os resultados da análise de correlação podem ser observados na Figura 19. Observa-se que ambos os testes utilizados indicaram que apenas um par de atributos poderia ser considerado altamente correlacionado pelos critérios usados: as amplitudes dos dois sensores. Esse é um resultado interessante, pois é esperado que a amplitude do sinal de correntes parasitas carregue boa parte das informações a serem obtidas na inspeção, sendo esse, por essa razão, um atributo com alta relevância em potencial.

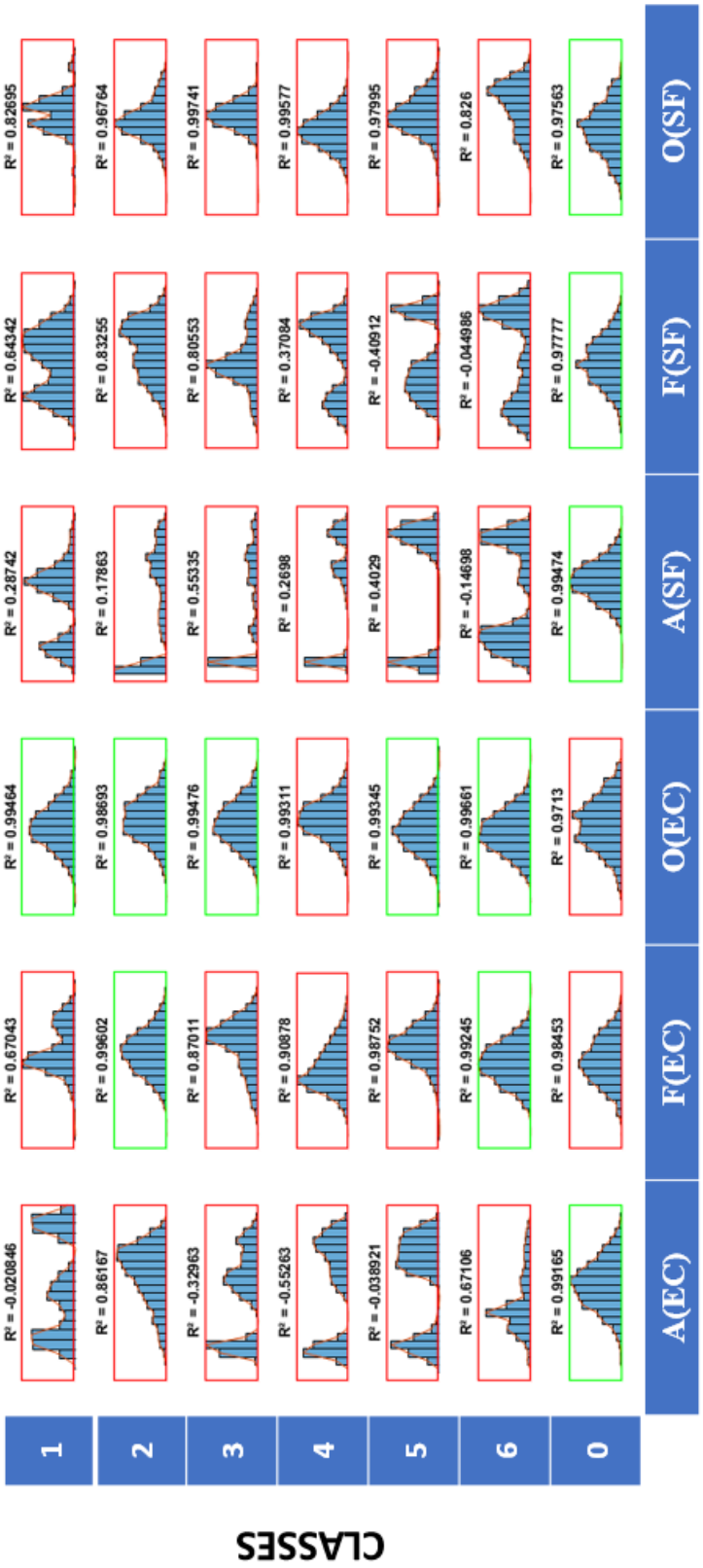


Figura 18: Resultados da análise de normalidade. Em verde, estão as distribuições consideradas normais pela análise, e em vermelho as consideradas não-normais. Também foram mostrados os coeficientes de determinação (R^2) de cada distribuição ao se tentar ajustar uma curva gaussiana, só como elemento ilustrativo.

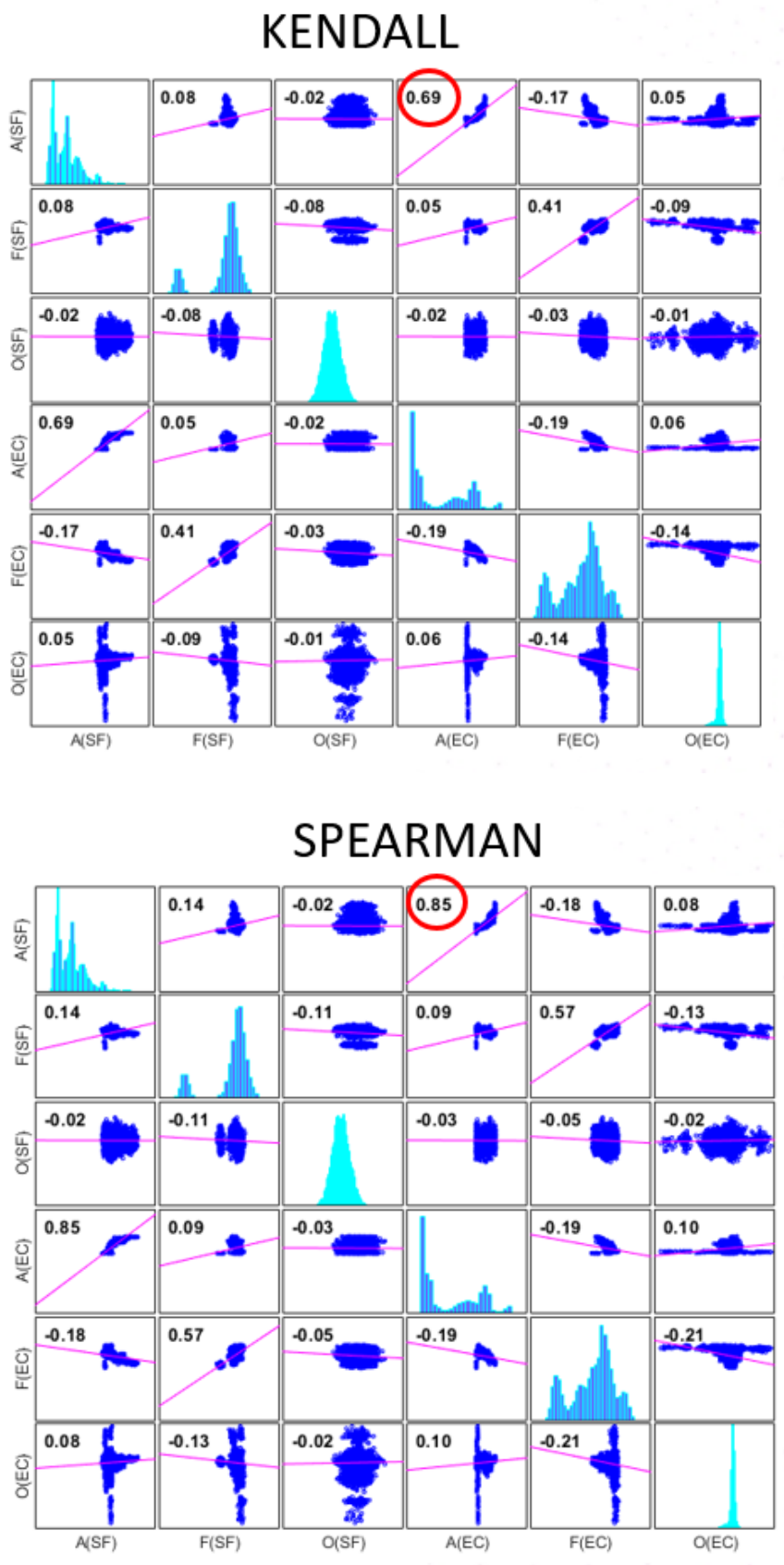


Figura 19: Resultados da análise de correlação. Os círculos vemelhos indicam os únicos pares de atributos que foram considerados altamente correlacioandos na análise.

Logo, esse resultado pode indicar que, possivelmente, o uso das amplitude dos dois sensores pode ser desnecessário, pela alta correlação que eles apresentam. Todavia, essa conclusão requer ser confirmada pelas análises posteriores, em especial pela análise de seleção de atributos.

Outro resultado que vale ser mencionado dessa análise é a correlação entre as fases de ambos os sensores, as quais mostraram uma correlação considerável (41% e 57% em cada teste), ainda que abaixo do critério estabelecido. Isso é um indício que, apesar da variabilidade dos dados das diferentes composições, o comportamento dos atributos em relação a eles é semelhante. Não obstante, as correlações entre os *offsets* foram baixas.

4.2.3. Análise de sobreposição

Os resultados da análise de sobreposição podem ser observados na Figura 20. Nessa figura, são mostrados em cada lacuna os atributos que não foram capazes de diferenciar o par de classes correspondentes.

Classes	0	1	2	3	4	5	6
0							
1				**O(SF)**	**O(SF)**		
2				**O(SF)**		**O(EC)**	**O(SF)**
3		**O(SF)**	**O(SF)**		**O(SF)** **F(SF)**		**O(SF)**
4		**O(SF)**					
5							
6			**O(SF)**	**O(SF)**			

Teste de Kolmogorov (TKS)		Teste de Wilcoxon (TPW)	

Figura 20: Resultados da análise de sobreposição.

Observa-se que os casos de sobreposição se deram para diversos pares de classes, porém principalmente para o atributo O(SF), tendo apenas um caso de sobreposição pelo atributo F(SF) e um pelo atributo O(EC). Esse resultado é um forte indicador do baixo desempenho de diferenciação das classes do problema pelo atributo O(SF) em comparação aos outros atributos. Consequentemente, espera-se uma baixa relevância desse atributo no desenvolvimento de modelos de classificação para dados da natureza do problema atual

4.2.4. Análise de seleção de atributos

Os resultados da análise de seleção de atributos podem ser observados na Figura 21. Esses resultados foram obtidos a partir do uso do algoritmo de filtragem RELIEF-F, que atribui um coeficiente de relevância de -1 a 1 para cada atributo do problema.

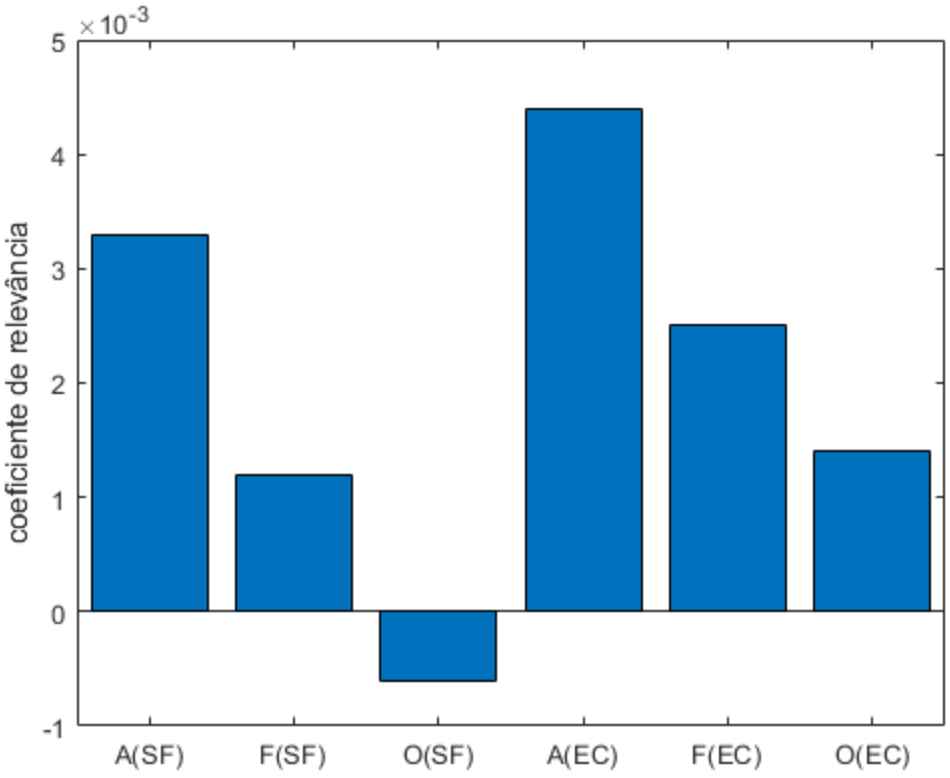


Figura 21: Resultados da análise de seleção de atributos.

Pelos resultados, é possível ordenar a relevância dos atributos por:

O(SF) < F(SF) < O(EC) < F(EC) < A(SF) < A(EC)

Essa ordenação mostra que, como esperado, ambos os atributos de amplitude apresentaram as maiores relevâncias. Isso os define, portanto, como os principais atributos do problema de classificação. Todavia, pelo resultado da análise da correlação ter indicado uma elevada correlação entre eles, a amplitude com menor relevância será selecionada como potencialmente irrelevante, ou seja, A(SF). Além disso, observa-se que o atributo O(SF) apresentou a menor relevância, com valor abaixo de 0, confirmando os resultados anteriores. Por essa razão, esse atributo também foi selecionado como potencialmente irrelevante. Ao fim da seleção, obtivemos um novo conjunto de atributos relevantes composto por A(EC), F(EC), O(EC) e F(SF). Ao fim da etapa de treinamento dos modelos, a influência do uso dos dois atributos potencialmente irrelevantes no desempenho dos modelos será testada.

4.3. Visualização dos dados

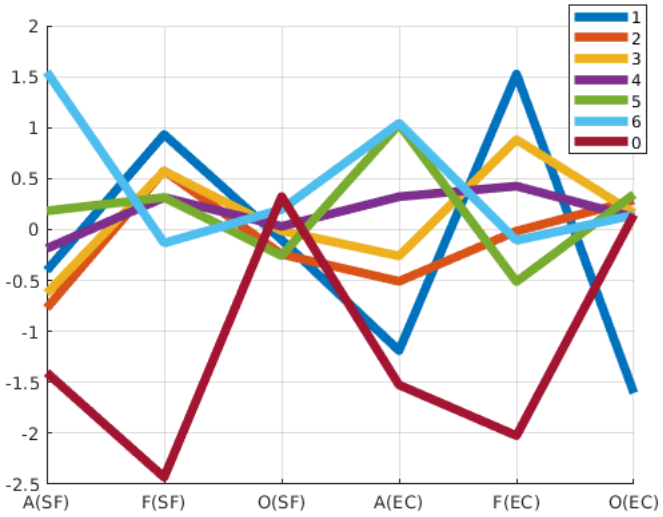
A visualização dos dados utilizando gráficos de coordenadas paralelas pode ser observado na Figura 22. Observa-se neles que os atributos avaliados como mais relevantes na análise anterior, de fato, visivelmente separam melhor as médias das distribuições estatísticas de cada classe, enquanto os atributos de menor relevância apresentam as médias em valores muito próximos. Essa baixa separação das médias de cada classe pelos atributos menos relevantes dificultam, portanto, a separação dessas classes pelo algoritmo de classificação. Esse resultado reforça e respalda os resultados numéricos das análises anteriores.

Além disso, pode-se observar também que as médias das distribuições estatísticas dessas classes são facilmente diferenciadas por certos atributos, como por exemplo:

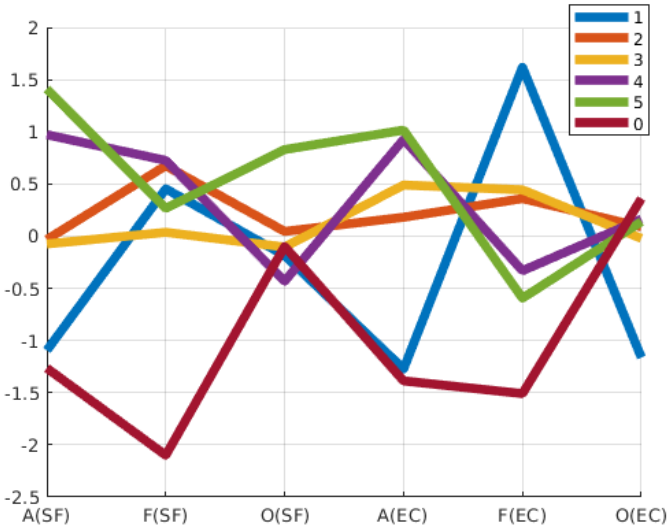
- A classe 0, que corresponde à sonda ao ar, tem sua média facilmente diferenciada das outras classes por ambas as fases;
- A classe 1, que corresponde ao estado I, tem sua média facilmente diferenciada das outras classes por F(EC) e O(EC);
- A classe 6, correspondente ao estado VI, tem sua média facilmente diferenciada das outras classes por A(SF);

Por se tratarem de classes associadas aos estados de evolução microestrutural extremos (I, VI e ar), é esperado que essas classes possam ser sua médias mais facilmente diferenciadas que as classes intermediárias.

(Conjunto A_P)



(Conjunto B_P)



(Conjunto AB_P)

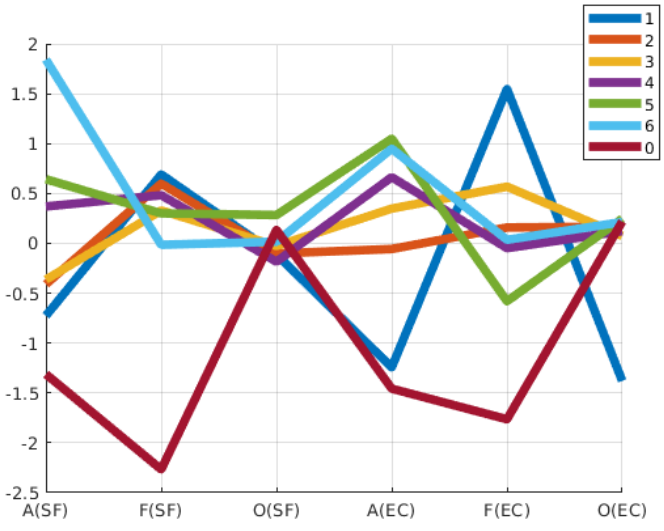


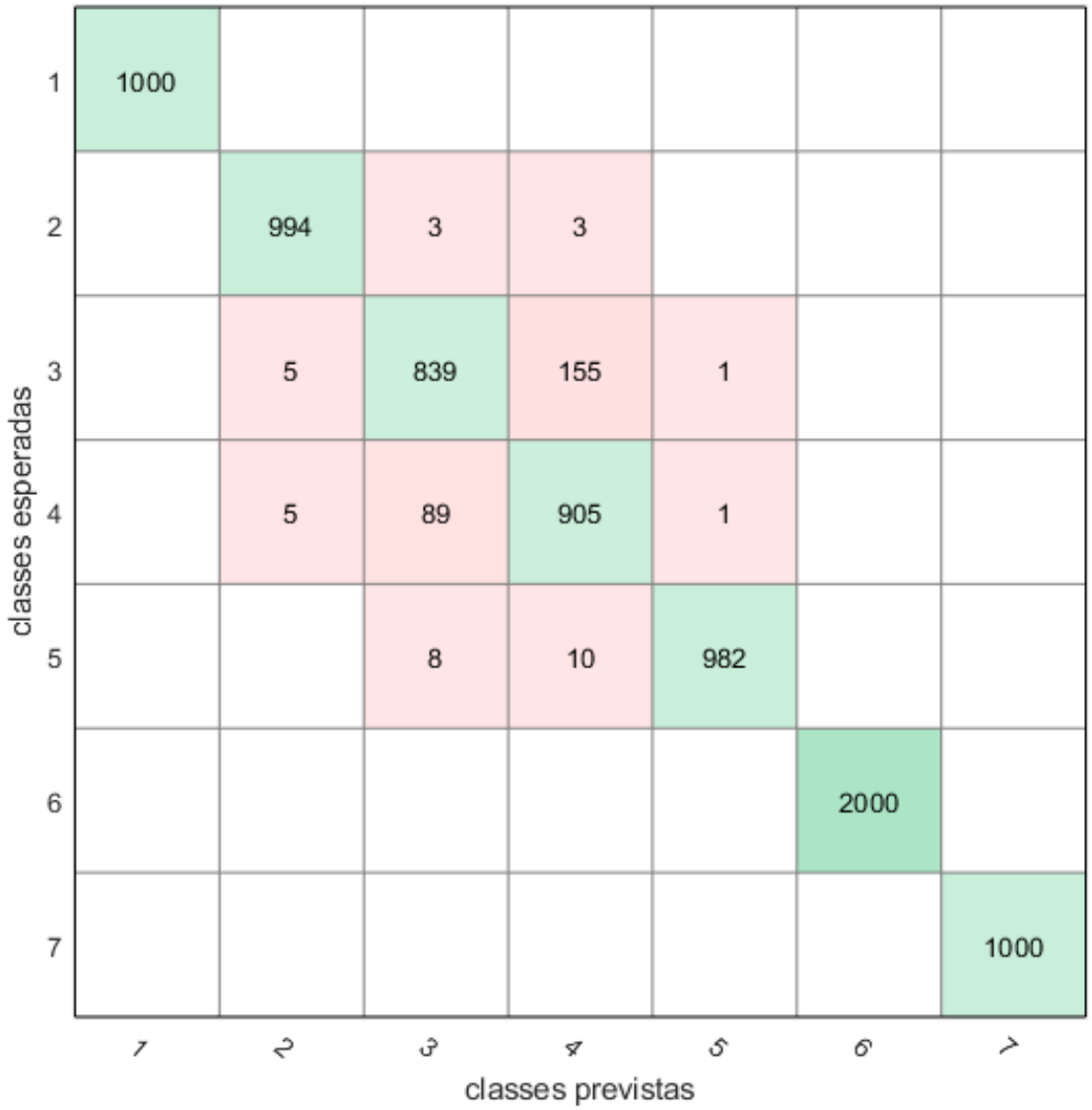
Figura 22: Visualização dos dados dos três conjuntos principais com gráficos de coordenadas paralelas. O eixo y dos três gráficos foi padrozinado para facilitar a visualização.

4.1. Treinamento dos modelos de classificação

Os resultados do treinamento dos três modelos MA, MB e MAB podem ser vistos nas Figuras 23, 24 e 25, respectivamente. Em cada figura estão expostas a matriz de confusão e as métricas numéricas calculadas para cada modelo.

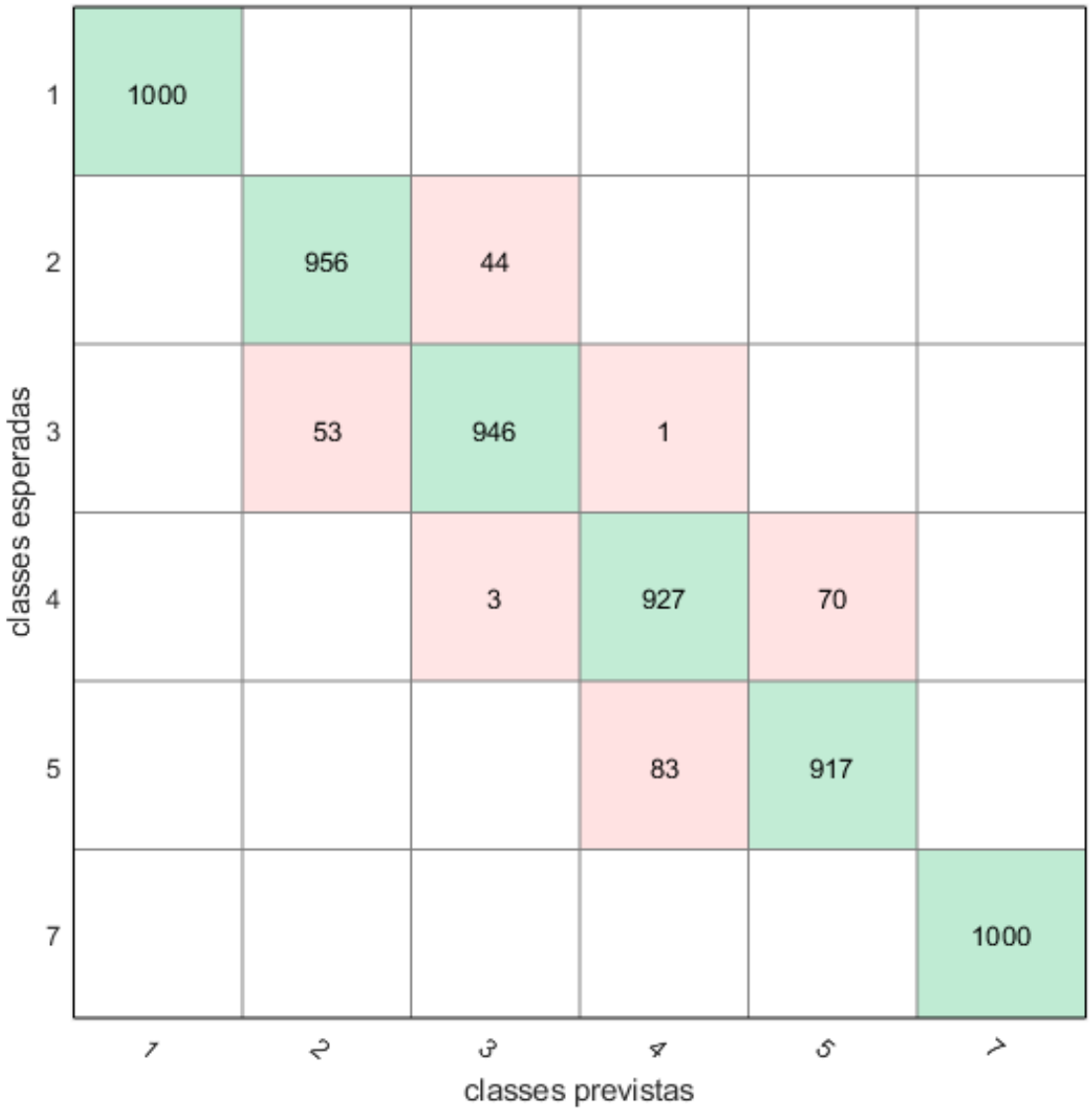
Em relação às classes, observa-se que, nos três modelos, os erros se concentram nos estados de evolução microestrutural intermediários, enquanto são mais baixos nos estados extremos (principalmente os estados I, II e IV). Esse resultado pode ser explicado pelo fato das características metalúrgicas dessas fases serem mais facilmente diferenciáveis de outros estados, o que gera uma separação mais clara das respostas magnéticas desses estados à inspeção por correntes parasitas. De fato, enquanto as transformações microestruturais do estado I para o estado II, do estado II para o III e do V para o VI são muito claras e geram variações consideráveis no volume, tamanho, e morfologia dos precipitados de cromo, as transformações entre os estados III, IV e V são mais sutis, envolvendo mais a distribuição e pequenas mudanças de tamanho desses. Isso também explica porque os erros de classificação entre estados de evolução não sucessivos (por exemplo, entre os estados II e IV), serem em geral mais baixos que entre estados sucessivos. Observa-se também que os três modelos cometeram erros nulos na classificação de dados associados a passagem da sonda ao ar.

Em relação ao desempenho geral de classificação de cada modelo, observa-se que as acurácias de validação dos modelos MA e MB são elevadas, estando todas acima de 95%. Isso significa que os modelos individuais obtém um elevado desempenho na classificação dos conjuntos de dados que foram usados em seus respectivos treinamentos, isto é, os grupos de amostras A e B, com composições químicas distintas. Todavia, suas acurácias de teste, isto é, as acurácias ao tentar classificar simultaneamente os dados provindos de ambos os grupos, são muito mais baixas, estando ambas abaixo de 75%. Esse resultado indica que ambos os modelos treinados com os dados de amostras de apenas uma composição química não foram adequados para a classificação de todos os dados. Por outro lado, para o modelo MAB, a acurácia de validação e teste (que para esse modelo são evidentemente iguais), é de 91.8%, o que mostra que a metodologia desenvolvida foi capaz de treinar um modelo de alto desempenho para um conjunto de dados com composições químicas variável. Além disso, observa-se que esse modelo obteve um coeficiente F1 de acertos acima de 85% para 6 das 7 classes do problema, apresentando alto erro apenas para o estado IV. Soluções para reduzir esse erro em específico serão focos de trabalhos futuros.



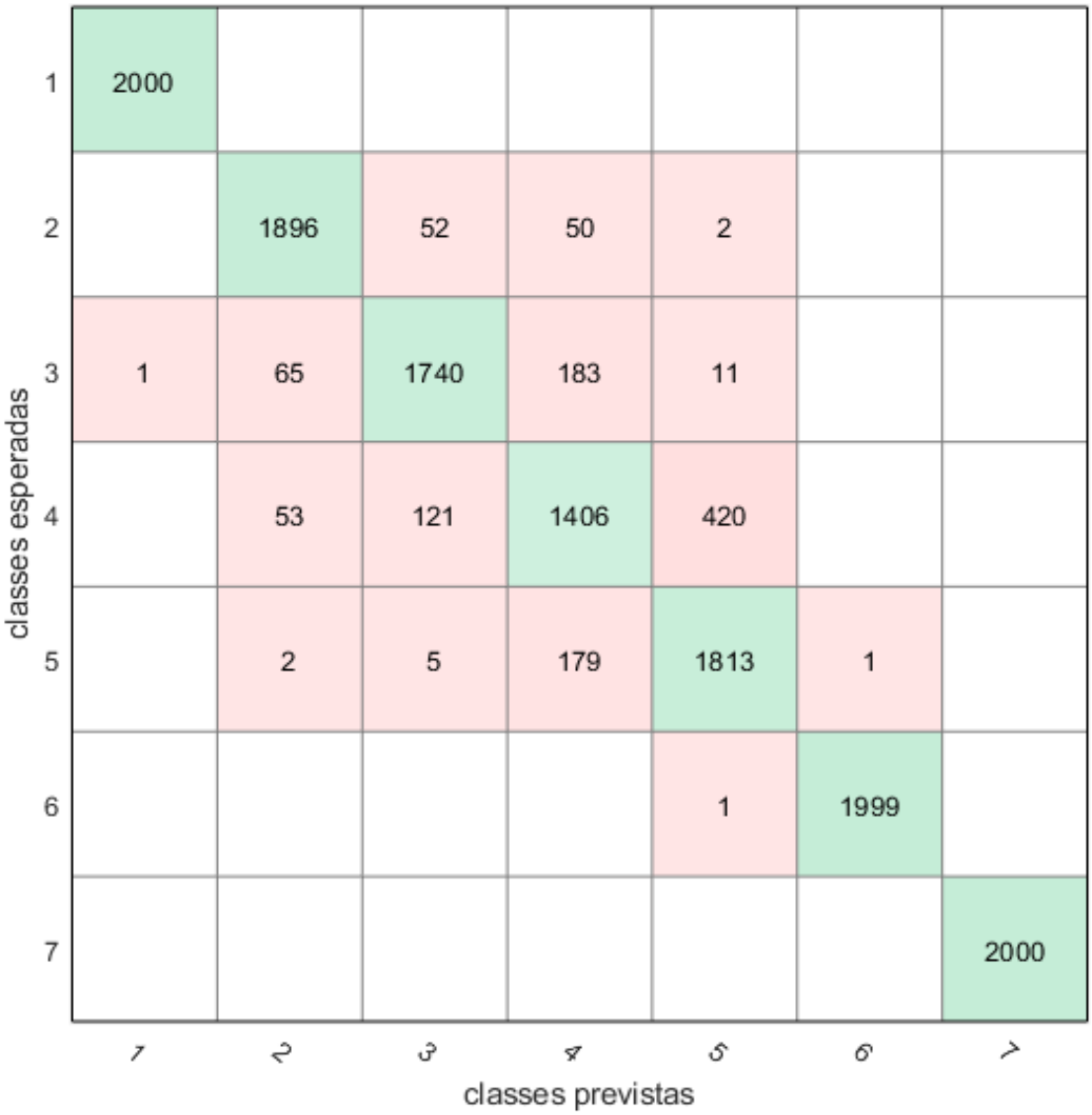
Modelo MA	Ar	I	II	III	IV	V	VI
Acurácia (valiação)	96.5%						
Precisão	100	100	99.0	89.0	84.0	99.9	100.0
Sensibilidade	100	100	99.0	84.0	91.0	98.0	100
Coeficiente F1	100	100	99.0	86.4	87.4	98.9	100
Acurácia (teste)	73.9%						

Figura 23: Resultados do treinamento do modelo MA. Acima, matriz de confusão e, abaixo, tabela de métricas numéricas calculadas em relação a cada estado de evolução microestrutural.



Modelo MB	Ar	I	II	III	IV	V	VI
Acurácia (validação)	95.8%						
Precisão	100	100	95.0	95.0	92.0	93.0	-
Sensibilidade	100	100	96.0	95.0	93.0	92.0	-
Coeficiente F1	100	100	95.5	95.0	92.5	92.5	-
Acurácia (teste)	61.6%						

Figura 24: Resultados do treinamento do modelo MB. Acima, matriz de confusão e, abaixo, tabela de métricas numéricas calculadas em relação a cada estado de evolução microestrutural.



Modelo MAB	Ar	I	II	III	IV	V	VI
Acurácia (validação)	91.8%						
Precisão	100	99.9	94.0	91.0	77.0	81.0	99.9
Sensibilidade	100	100	94.8	87.0	70.3	90.6	99.9
Coeficiente F1	100	99.9	94.5	88.9	73.3	85.7	99.9
Acurácia (teste)	91.8%						

Figura 25: Resultados do treinamento do modelo MAB. Acima, matriz de confusão e, abaixo, tabela de métricas numéricas calculadas em relação a cada estado de evolução microestrutural.

4.1. Avaliação da seleção de atributos

Os resultados da avaliação de seleção de atributos podem ser vistos na Tabela 15. Na tabela, são mostrados os valores de acurácia de validação para os conjuntos de atributos testados utilizando a técnica de validação cruzada com k = 10. Desses testes, observa-se que para os três modelos a exclusão do atributo O(SF) gerou um aumento de erro de menos de 0.5% nos treinamentos. Esse resultado conclui que, como previamente indicado pela análise de seleção de atributos, esse atributo de fato apresenta uma relevância desprezível para o problema de classificação. Por essa razão, ele pode ser excluído do conjunto ideal de atributos.

Tabela 15: Resultados da análise de seleção de atributos.

	MA	MB	MAB
Com todos os atributos	96.5%	95.6%	91.8%
Sem o atributo O(SF)	96.5%	94.5%	91.3%
Sem o atributo A(SF)	94.7%	94.3%	86.6%
Sem os atributos O(SF) e A(SF)	94.7%	93.2%	86.2%

Por outro lado, a exclusão do atributo A(SF) levou a uma redução considerável da acurácia de classificação dos modelos, em especial o modelo único MAB. Esse resultado mostra que, apesar da análise estatística mostrar uma alta correlação desse atributo com o A(EC), sua relevância permanece alta para a classificação, reforçando o resultado da análise de seleção de atributos. Por essa razão, esse atributo não pode ser excluído do conjunto ideal de atributos.

CONCLUSÃO

O estudo realizado demonstrou que a metodologia proposta foi bem sucedida no treinamento de um modelo de inteligência artificial capaz de classificar com acurácia de 91.8% os estados de evolução microestrutural de amostras de tubos de reforma a vapor de aço HP com duas composições químicas distintas.

As análises de pré-processamento realizadas permitiram preparar satisfatoriamente os dados provenientes dos dois grupos de amostras para o processo de treinamento, além de permitir a compreensão com mais clareza das suas características e variações estatísticas. Em especial, a análise de seleção de atributos permitiu indicar aqueles atributos que poderiam ter uma relevância desprezível na classificação, podendo, portanto, serem excluídos dos modelos ao ter essa baixa relevância confirmada.

Com o processo de treinamento, e posteriormente de validação e teste, dois modelos que foram treinados cada um com dados de tubos com apenas uma das duas composições químicas tiveram seu desempenho de classificação comparados ao modelo único. Essa análise mostrou que, apesar dos modelos individuais terem um desempenho mais elevado classificando os dados de cada composição química, na classificação simultânea dos dados das duas composições o modelo único teve um desempenho muito mais elevado. Além disso, mostrou-se que o modelo único obteve um coeficiente F1 de acertos acima de 85% para 6 das 7 classes do problema, apresentando alto erro apenas para o estado IV.

Por fim, avaliou-se o aumento de erro pela exclusão dos atributos indicados como pouco relevantes pela análise de seleção de atributos. Essa avaliação levou a determinar a exclusão do atributo O(SF), uma vez que sua retirada tenha baixo impacto no desempenho dos três modelos treinados.

REFERÊNCIAS

Amid, E., Aghdam, S. R. and Amindavar, H. 2012. Enhanced Performance for Support Vector Machines as Multiclass Classifiers in Steel Surface Defect Detection. *International Journal of Electrical and Computer Engineering.* 2012. Vol. 6, 7.

Arenas, M. P. Caracterização magnética da microestrutura de diferentes estados de envelhecimento de uma liga austenítica HP modificada ao Nb-Ti. Tese de doutorado - COPPE/UFRJ, 2017.

Arenas, M. P., Caracterização por ensaios magnéticos não destrutivos de aços HP provenientes de fornos de reforma, Dissertação de mestrado - PEMM/UFRJ, 2013.

ASTM-A608/A608M; Standard specification for centrifugally cast iron-chromium nickel high-alloy tubing for pressure application at high temperatures; Stand. ASTM, vol. 01, pp. 1–6 (2006).

Barbabela, G. D., *Estabilidade Estrutural de Aços Resistentes ao Calor do Tipo HP com Adição de Nióbio*; Tese de D.Sc.; COPPE/ UFRJ, Rio de Janeiro, RJ, Brasil, 1990.

Bonaccorsi, L. Damage analysis in Fe-Cr-Ni centrifugally cast alloy tubes for reforming furnaces. Engineering Failure Analysis, vol. 36, pp. 65–74 (2014).

Boslaugh, Sarah. Statistics in a Nutshell. Sebastopol, CA: O'Reilly Media, 2012. Print.

Camerini, C. G. Sistema de inspeção para detecção de trincas de fadiga em juntas soldadas de dutos metalurgicamente cladeados. Dissertação de doutorado - COPPE/UFRJ, 2018.

Cho, D. H. and Lee, S. L. 2013. Defect Classification Using Machine Learning Techniques for Flat Display Panels. *Applied Mechanics and Materials.* 2013. Vols. 365-366, 720–724. 10.4028/www.scientific.net/amm.365-366.720.

Dangeti, P. (2017). Statistics for Machine Learning. Birmingham: Packt Publishing.

Dutra, C. C. G., Análise de dados de inspeção não destrutiva em tubos de reforma a vapor utilizando treinamento de máquina. Projeto de Graduação, Engenharia de Petróleo, Escola Politécnica, Universidade Federal do Rio de Janeiro (2019).

LINK, Rainer; RIESS, Nathanael. (2018). NDT 4.0 - Overall Significance and Implications to NDT. 12th European Conference on Non-Destructive Testing (ECNDT 2018), Gothenburg 2018, June 11-15 (ECNDT 2018).

Lorena, A. C., De Carvalho A. C. P. L. F. Introdução à Máquinas de Vetores de Suporte, Relatórios Técnicos do ICMC. (2003) 66.

Marques, P. E. M. B. et al, in *XXII Encontro Nacional de Modelagem Computacional* (2019), p. 177 a 179.

Martins, M. V. M. Correlação entre os sinais de correntes parasitas e o nível de dano em tubos de aço HK-40, Dissertação de mestrado - PEMM/UFRJ, 2000.

MathWorks (2019). *Statistics and Machine Learning Toolbox: User's Guide (R2019a)*. Acessado em 29 de janeiro de 2020 em www.mathworks.com/help/pdf_doc/stats/stats_ug.pdf.

NDT.ED - Non Destructive Testing Education [Internet]. Disponível em: www.ndt-ed.org. Acessado em 09 de Janeiro de 2020.

NDT.net, “The Largest Open Access Portal of Nondestructive Testing” [Internet]. Disponível em: www.ndt.net. Acessado em 13/01/2020.

Queiroz, F. S. Caracterização microestrutural em função da temperatura de serviço de tubos de aços HP-modificados ao Nb e Ti envelhecidos em fornos de reforma a vapor, Dissertação de mestrado - PEMM/UFRJ, 2017.

Queiroz, F. S.; Influência de surto de temperatura sobre o comportamento dos tubos de fornos de reforma fundidos por centrifugação em liga hp-40 modificada ao nióbio. Projeto de Graduação, Engenharia Metalúrgica, Escola Politécnica, Universidade Federal do Rio de Janeiro (2012).

Rezende, S.O. (ed). Sistemas Inteligentes, Manole, 2003, ISBN 85-204-1683-7.

Santos, Rodrigo Magalhães Mota dos. Técnicas de aprendizagem de máquina utilizadas na previsão de desempenho acadêmico – 2016. Dissertação (Mestrado em Educação) – Universidade Federal dos Vales do Jequitinhonha e Mucuri, 2016. Programa de Pós-Graduação em Educação.

Schulenburg, Lennart (2019) NDT 4.0 – Opportunity or threat? [internet]. Acessado em 04/02/2020. Disponível em: https://www.linkedin.com/pulse/ndt-40-opportunity-threat-lennart-schulenburg/.

Silveira, T.L. and May, I.L. Reformer furnaces: materials, damage mechanisms and assessment. The Arabian Journal for Science and Engineering, 2006, Vol. 31, pp. 99-119.

Spruyt, Vincent (2014). The Curse of Dimensionality in classification [internet]. Acessado em 30 de Janeiro de 2020. Disponível em: https://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/.

Srinidhi, S. (2018), "Different types of Validations in Machine Learning (Cross Validation)", Acessado em 08/08/2019. https://blog.contactsunny.com/data-science/different-types-of-validations-in-machine-learning-cross-validation.

Swaminathan, J., et al. Failure analysis and remaining life assessment of service exposed primary reformer heater. Engineering Failure Analysis, 2008, Vol. 15, pp. 311-331.

Ting, Kai Ming (2011). Encyclopedia of machine learning. Springer. ISBN 978-0-387-30164-8.

Tomiska, J. Vienna. The system Fe–Ni–Cr: revision of the thermodynamic description. Elsevier, 2004, Journal of Alloys and Compounds, pp. 176-187.

Urbanowicz, Ryan & Meeker, Melissa & LaCava, William & Olson, Randal & Moore, Jason. (2017). Relief-Based Feature Selection: Introduction and Review. Journal of Biomedical Informatics. 85. 10.1016/j.jbi.2018.07.014.

Victor Aviation TM. 2019. Eddy Current Inspection. Victor Aviation TM website. [Online] Victor Aviation TM, 2019. [Cited: 06 27, 2019.] http://www.victor-aviation.com/Eddy-Current-Inspection.php.

Vrana, Johannes. (2019). NDE 4.0: The Fourth Revolution in Non-Destructive Evaluation: Digital Twin, Semantics, Interfaces, Networking, Feedback, New Markets and Integration into the Industrial Internet of Things. 10.13140/RG.2.2.17635.50720.

Wang, Y. and Guo, H. 2014. Application of Support Vector Machine in Weld Defect Detection and Recognition of X-ray Images. *Computer Aided Drafting, Design and Manufacturing.* 2014. Vol. 24, 3.

Xristica (2019) What is the difference between feature extraction and feature selection? [internet]. Acessado em 05/02/2020. Disponível em: https://quantdare.com/what-is-the-difference-between-feature-extraction-and-feature-selection/

Yadav, S.; Shukla, S. (2016) "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram, pp. 78-83.